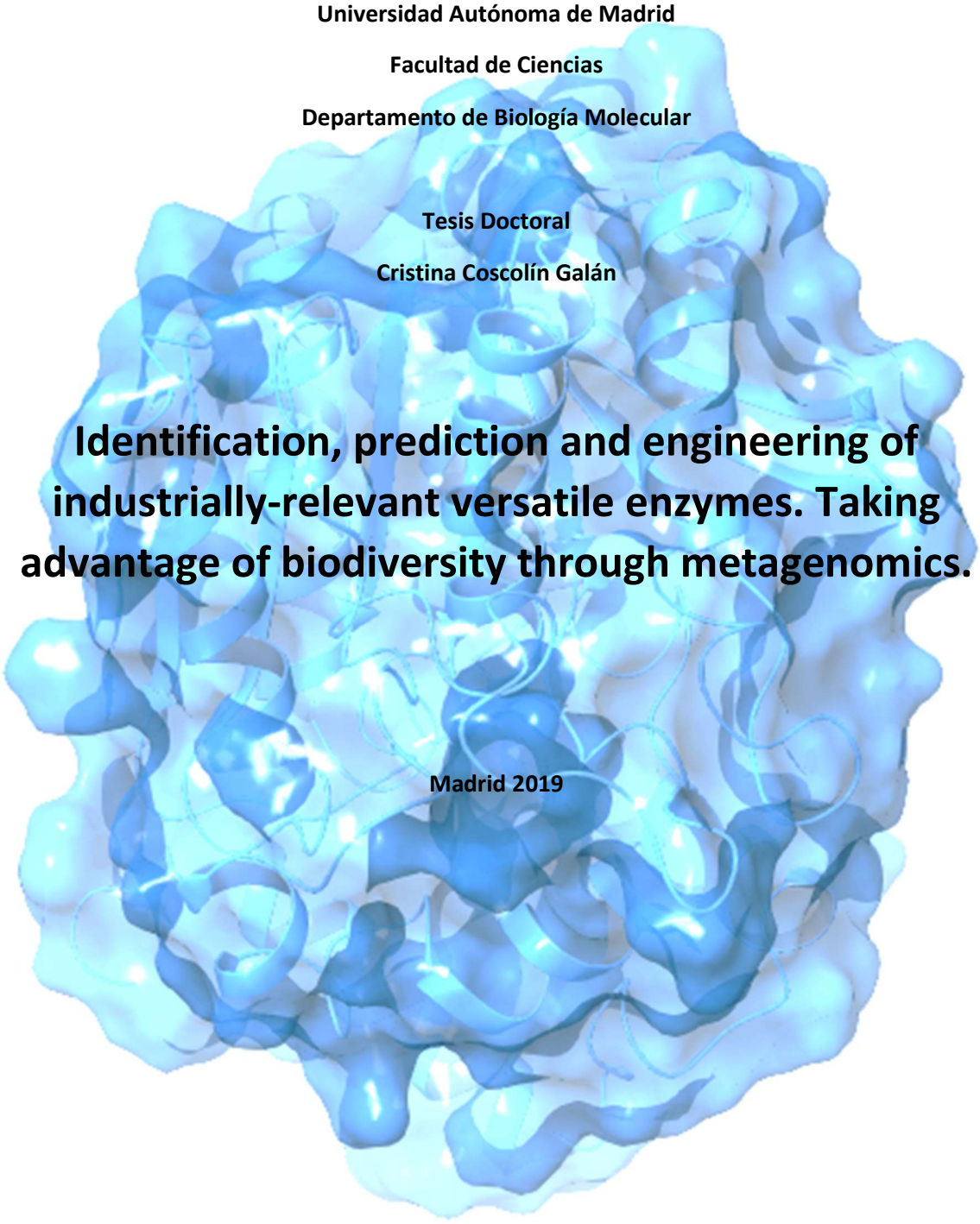
Universidad Autónoma de Madrid

Facultad de Ciencias

Departamento de Biología Molecular

Tesis Doctoral

Cristina Coscolín Galán



Identification, prediction and engineering of industrially-relevant versatile enzymes. Taking advantage of biodiversity through metagenomics.

Madrid 2019

Universidad Autónoma de Madrid

Facultad de Ciencias

Departamento de Biología Molecular

Identification, prediction and engineering of industrially-relevant versatile enzymes. Taking advantage of biodiversity through metagenomics.

Autora: Cristina Coscolín Galán. Licenciada en Biología

Memoria para optar al grado de Doctora por la Universidad Autónoma de Madrid

Director: Doctor Manuel Ferrer Martínez

Tutora: Profesora Elena Bogónez Peláez

Tesis Doctoral realizada en el Instituto de Catálisis y Petroleoquímica

perteneciente al Consejo Superior de Investigaciones Científicas

Madrid, 2019

MANUEL FERRER MARTÍNEZ, DR. EN CIENCIAS QUÍMICAS, INVESTIGADOR CIENTÍFICO DEL CSIC

CERTIFICA: que el presente trabajo "Identification, prediction and engineering of industrially-relevant versatile enzymes. Taking advantage of biodiversity through metagenomics", que constituye la Memoria que presenta la Licenciada en Biología por la Universidad de Alcalá, Cristina Coscolín Galán, para optar al grado de Doctora, ha sido realizado bajo su dirección en el Departamento de Biocatálisis del Instituto de Catálisis y Petroleoquímica del CSIC, Campus de Excelencia Internacional UAM + CSIC, Madrid.

Y para que conste, firma el presente certificado en Madrid, Diciembre de 2019.

Dr. Manuel Ferrer Martínez

A mis abuelos

"Haz primero lo que tú quieras,

que no llegue el día de mañana y te arrepientas

de no haberlo intentado"

Trinidad Galán Sánchez, 2007

ACKNOWLEDGEMENTS/AGRADECIMIENTOS

En primer lugar, quiero dar las gracias al Ministerio de Economía y Competitividad, Ministerio de Ciencia, Innovación y Universidades, a la Agencia Estatal de Investigación (AEI) y al Fondo Europeo de Desarrollo Regional (FEDER) por la Beca FPI (BES-2015-073829) y la financiación de los proyectos a los cuales está asociada dicha beca (BIO2014-54494-R, BIO2017-85522-R). También a los miembros de los consorcios INMARE y METACAT, que han puesto a mi disposición algunas de las muestras y enzimas caracterizadas en la presente Tesis.

Me gustaría agradecer a mi director de tesis, el doctor Manuel Ferrer. Primero, que me diera la oportunidad de realizar una tesis doctoral y, después, la infinita paciencia que ha tenido siempre conmigo y con mis despistes, que no son pocos ni mucho menos; el cariño y el respeto que me ha mostrado siempre, todo lo que me ha enseñado en estos años, y por hacer del laboratorio un segundo hogar para mí.

Gracias a mis compañeros de laboratorio, los que están y los que se han marchado, por compartir conmigo estos años, por las risas, por enseñarme y por ayudarme tanto. En particular, gracias a Rafa por sus maravillosas gráficas y por las charlas de tatuajes y videojuegos; gracias a Pepe por los buenos momentos en las comidas durante las reuniones de proyecto; gracias a Natalia por el breve, pero divertidísimo verano que pasó en el laboratorio; gracias a Laura por hacer inventario del "labo" y por recomendarme buenos *animes*; gracias a David por encargarse del trabajo "sucio" y por animarme siempre; gracias a Bartolo por hacer del despacho un sitio siempre alegre, por llevarme de excursión y por compartir conmigo la guerra contra los "portazos"; gracias a Sandra por ser mi compi de manifestaciones y de restaurantes veganos, y por darle "estilasso" al grupo; gracias a mi Moni, por enseñarme tantísimas cosas, por ser un ejemplo a seguir de pulcritud científica, de solidaridad, de paciencia, por ese corazón que no te cabe en el pecho…, mil veces gracias.

Gracias a todo el personal del ICP y de la Universidad Autónoma de Madrid. En particular, al director y a los compañeros de administración, de mantenimiento y de limpieza del ICP por toda su ayuda y apoyo. A Elena, mi tutora de la universidad por su supervisión y por agilizar siempre los tramites, y a Magda por resolverme pacientemente todas mis dudas sobre la tesis. Gracias Carmen por tirarme las latas de Coca-Cola medio vacías, por consolarme tantas frías mañanas, por ser tan buena persona y por ser mi amiga. Gracias a los "Alcaldes", a Patch que siempre tiene la solución a nuestros problemas, a Paty con la que comparto gusanadas y mapaches, a Xavi que me ayudó con los papeles de la mención internacional, a Berni por tantos momentos divertidos y por su ayuda con los papeles de la estancia, a Iván por las risas en el Biotec, a Isa V por los viajes y las "risis", y por supuesto, a Javi por las incontables veces que me ha apoyado, ayudado, animado y escuchado, gracias de todo corazón mi amigo. Muchas gracias a las "Kikas", a Lopa por ser tan *cool* y hacer las mejores *playlists*, a Lu por distraer a los comerciales, a Bárbara que siempre me invita a la clase de metagenómica, y a Fadia por el vicks vaporub y por las risas de última hora de la tarde; a Kiko, Antonio y Joselu, por sus amenas charlas del café de las 11, y gracias a David "el niño" por tenerme al día de las modas modernas.

I would like to thank professor Olivier Lichtarge for letting me do a short-stay in his Lab at the Baylor College of Medicine in Houston. Thank you, Brigitta, for helping me with all the bureaucracy. Thank you, Panos, for being such a patient instructor and for teaching me so many things about Evolutionary Trace; thanks to my room-mates Hunter, Christina and Gary for making me feel so comfortable so far from home; also thanks to Kico and Boo for being so kind and lovable. Many thanks to my friends Eunna and Lynn, for all the fun and all the great food we shared, I can't wait for you to come visit me. Thank y'all guys!

También querría agradecer a la profesora María Fernández Lobato por haberme dejado ir al CBM a aprender a expresar en *Pichia*; a Zoran, que se encargó de enseñarme y con quien siempre me divierto un montón, a Peter por ayudarme a terminar los chupitos en las fiestas, y por supuesto a Martín, "mi ídolo", que siempre tiene alguna sabia lección de la vida que enseñarme, porque siempre me hace reír hasta que se me saltan las lágrimas y por ser tan cariñoso sin importarle que yo sea una rancia.

Así mismo, al profesor Víctor Guallar por acogerme en el Centro de Supercomputación de Barcelona (BSC) una semanita, y a Gerard que fue quien me enseñó las primeras cosas de bioinformática y quien siempre me ayuda cuando meto la pata con las simulaciones; gracias a todos los chicos del BSC con los que me divertí tantísimo, y a Ferran por ser mi compi de estancia transoceánica y por su ayuda con los comandos malvados.

Gracias a la doctora Julia Sanz por dejarme pasar de vez en cuando por el laboratorio de cristalografía y a Isa (Isita de mi corazón) por sacarme cristales como churros.

Quiero mencionar también al profesor Ángel Martín por permitirme comenzar mis andanzas en la ciencia en el grupo de biotecnología para la biomasa lignocelulósica, donde realicé el trabajo de fin de master y pasé dos años increíbles. Gracias a toda mi gente del CIB (sois tantos que alguno se va a quedar por el camino, lo siento de antemano): a Javi, Mª Jesús, Susana, Ali, Marta, Aitor, Vero, Davi, Mariu, Jorge, Felipe, Manu, Isita P, Jezú, Mario, María, Meme y Juan Carro; gracias por los partidos de pádel, las cañas en el Ondiñas, las fiestas de la primavera, las casas rurales y las miles de aventuras que pasamos todos juntos. Haré una mención especial para mis *ladies* del aquelarre: a Ana por traerme la gracia de Zaragoza más cerca, a Lidem por todas las risas, por los memes, por las recetas "absurder", por los rescates gatunos y por reintroducirme en la "secta" de los patines, y a mi Lolita, por ser la primera en enseñarme a trabajar en el labo, la más paciente, la más dulce y la más buena.

Un agradecimiento especial se merecen mis queridísimos "muebles". Gracias Joan por tu destreza y tu genialidad; hacer música contigo es fácil para cualquiera a la par que divertido. Gracias Carlos por compartir tu increíble talento una temporada conmigo, por tu apoyo y por tu cariño. Gracias Iván, mi leal compañero por enseñarme a no conformarme con lo que hay y a luchar por mejorar siempre.

A mis compañeros de la carrera, a Javi y Eli, por todos los buenos momentos en la facultad y porque, aunque ahora no nos vemos todo lo que me gustaría, siempre seréis mi "biofamily". Gracias a Edu, por tu amistad siempre constante, por las risas, por las pipas, y por tu apoyo incondicional.

Gracias a mis amigos de Guadalajara, por los festivales, las noches de fiesta, los dardos y la diversión sin fin, y en particular a Antonio por compartir su amistad conmigo todos estos años, por su apoyo en los momentos malos y por su compañía en los buenos. También a Aitor por las tardes de Play y series, las ranoaventuras, las gyozas y las sales de frutas especiales.

Gracias al camionero que me sacó del coche aquella lluviosa mañana de marzo, y a "la Gorda" que me ayudó a recuperarme de aquel duro golpe.

Gracias a mis amigos de Madrid Patina, a mi profe Dave, con el que he mejorado tanto, a mis "rollergorders" con los que un día malo siempre puede acabar en risas y pizzas, a Meju por sus ánimos y por ser mi compi de ejercicios, a Pañi que hace siempre que cualquier actividad en su compañía sea agradable, a Ross por su desparpajo y por ser tan divertida, a Heleneras por ser tan buena gente y por enseñarnos a seguir adelante por muy dura que sea la lesión, a Roci por compartir los momentos "haters" contra la gente tonta y el reggaetón, y a Cris por no salir del grupo aunque seamos muy pesados y por traernos manolitos el último día de clase.

Quiero darle las gracias a mi amiga Noa por ayudarme tantísimo con esta tesis y con todo lo demás, por las tardes de cine y gofres, por la diversión acuática, ya sea en playa o en piscina, por las partidas al Spyro, por el bádminton, por el pádel, por compartir comida rica a medio día y Coca-Colas en las tardes, por escucharme, por consolarme y por hacerme reír. También a la mamá de Noa, a David "el cigüeño", a Isaac "el demonio", a Tupi, a Hachiko y a Puchí que me hacen sentir siempre en casa cuando visito Galicia.

No querría dejar de agradecer a mi "amiga a través del tiempo y el espacio", Lorea, porque, aunque pase el tiempo y estemos lejos, siempre estás ahí con tus sabios consejos, con tus memes divertidos, con tus palabras de ánimo y cariño, con tus *papers* geniales, y por tu increíble e inquebrantable amistad.

Por último, quiero darle las gracias a mi familia, que siempre me ha apoyado en mis decisiones y siempre ha estado ahí a mi lado. Primero, a la "no humana": mis perros, mis gatos, mi pájaro, mi hurona, mis ranas, mi araña..., los que os marchasteis y los que estáis aquí; gracias por enseñarme que todas las vidas importan, a no ignorar el dolor de los demás, y por vuestro cariño sin límites. Y después a la "humana", especialmente a mi abuela, que no ha podido verme terminar la tesis, pero que sin ella nada hubiera comenzado, gracias por enseñarme a ser fuerte y a no dejarme amedrentar por nadie; a mi abuelito, que nunca me ha dado otra cosa que no sea cariño y cuidados; a mi prima Patricia por las innumerables risas que hemos compartido y por ser siempre un ejemplo de mujer valiente para mí; a Javo por cuidar de Patito; a mi hermano Ramón por su incomparable sentido del humor y por ser siempre tan bueno y cariñoso; a Silvi por aguantar a Ramón; a mi padre Juan por enseñarme a pelear cuando algo no es justo y porque, aunque sé que a veces no me entiende, nunca me ha dejado de defender; a mi mami Trini "el morspo", por inculcarme todos los valores que hacen de mí la mujer que soy hoy, por animarme siempre a conseguir mis metas, por creer en mí más que nadie en el mundo, por protegerme, por recogerme del suelo cuando me he caído, y estar a mi lado todos estos años…, nunca te lo podré agradecer lo suficiente, mamá.

Gracias a todos y a todas, de corazón.

INDEX

ABSTRACT

Enzymes are nature's catalysts that can help industry, as they aid to making industrial processes and relevant products such as those in food, agricultural, cosmetic, and pharmaceutical sectors, faster, cleaner, safer, cheaper, and more sustainable. Their use contributes to solving global challenges such as overpopulation, diminishing natural resources, pollution and human health challenges, which in turn demands new enzymes with novel and/or improved properties, structures and activities. Finding new enzymes for academic and industrial settings is a long and complex process, but due to extensive efforts of the scientific community, this time frame has been significantly reduced: nowadays, it takes less than three years to provide industry with new enzyme solutions. Most enzymes are isolated from microbes and recent technological advances mean that it is relatively easy to produce new enzymes. The real problem is that only a very small percentage of new enzymes are useful in industrial processes. This was this Thesis's challenge, to improve the value of new enzymes and to help understanding, predicting and engineering their properties, to find candidates of interest for industry. First, this Thesis set out to find new enzymes from phylogenetically diverse uncultured microbial communities as well as cultured microbes, that develop in multiple environments, including some of the most extreme habitats on the planet. This guarantees high enzymatic diversity and that the recovered enzymes may be already adapted to work in multiple environmental conditions of pH, temperature, salinity, and nutrient diversity and concentration, to cite some. Second, this Thesis focuses its attention on finding and improving the value of new enzymes for industry, by targeting so-called "substrate promiscuous" enzymes, that are capable of accepting many substrates and therefore are usable in more than one industrial process. By focusing on serine ester-hydrolases and class III ω-transaminases, and by applying and designing a number of screening tools and high throughput technologies, we have created one of the largest collection of such enzymes in a laboratory, accounting for a total of 155 (145 and 10, respectively). By their extensive characterization with circa 130 different substrates, the substrate specificity and selectivity of each enzyme candidate was evaluated, and those with prominent features identified. With the help of crystal structures and computational tools, markers of "substrate promiscuity" and selectivity were identified, through which this Thesis was able to fast-track the identification of industrially versatile ester-hydrolases and class III ω-transaminases; they have better properties, namely, higher and uncommon substrate specificity and selectivity, compared to best commercial enzymes. Further, we demonstrated that it is possible to transform a "substrate promiscuous" but non stereo-specific enzyme into a biocatalyst with broad substrate range while showing chiral selectivity through supramolecular and protein engineering tools. Major outcomes of the present Thesis include: I) the generation of robust meta-data demonstrating that biodiversity in the environment provides an invaluable source of novel enzymes; II) the accumulation of data demonstrating that "substrate promiscuous" enzymes accepting a wide range of different substrates can be identified, and are more abundant than expected; III) the demonstration that the "substrate promiscuity" level of certain classes of enzymes can be predicted from the analysis of their sequence, 3D model and/or structure, without the need of previous cloning, expression and characterization; IV) reinforcing the idea that broad substrate range is associated with low enantio-selectivity, but that this can be solved by supramolecular engineering and by applying to an enzyme flexibility constraints through different immobilization strategies; V) the demonstration that through discovering novel binding sites where extra active sites can be introduced, it is possible to generate an enzyme with two functional reactive groups, to improve the competitiveness and catalytic opportunities of enzymatic catalysts.

RESUMEN

Las enzimas son catalizadores naturales que pueden ser de utilidad en el sector industrial, ya que contribuyen a hacer que algunos procesos, como aquellos relacionados con empresas de los sectores alimentación, agrícola, cosmético y farmacéutico, sean más rápidos, más limpios, más seguros, más baratos y más sostenibles. Su uso puede contrarrestar problemas globales derivados de la sobrepoblación como la disminución de recursos naturales, la polución y los problemas de salud, que a su vez demandan enzimas con nuevas y/o mejores propiedades. Encontrar nuevas enzimas es un proceso largo y complejo, si bien el tiempo necesario para su descubrimiento se ha reducido significativamente: actualmente, es posible proporcionar a la industria nuevas soluciones enzimáticas en menos de tres años. La mayoría de las enzimas son aisladas de microorganismos y los últimos avances tecnológicos hacen que sea relativamente fácil producir nuevas enzimas. El problema estriba en que solo un pequeño porcentaje de estas nuevas enzimas es útil para los procesos industriales. Este ha sido uno de los retos de esta Tesis, aumentar el valor de nuevas enzimas y ayudar a comprender, predecir y modificar sus propiedades, para encontrar candidatos de interés para la industria. Primero, nos propusimos encontrar nuevas enzimas a partir de un amplio muestrario de comunidades microbianas y microorganismos filogenéticamente diversos, que se desarrollan en ambientes heterogéneos, incluyendo algunos de los más extremos del planeta. Esto garantiza una amplia diversidad enzimática y que las enzimas aisladas se encuentren adaptadas a trabajar en múltiples condiciones de pH, temperatura, salinidad y diversidad y concentración de nutrientes, por citar algunas de las más relevantes. Segundo, esta Tesis centra su atención en encontrar y aumentar el valor de nuevas enzimas para la industria, especialmente en las llamadas enzimas de amplia "promiscuidad de sustrato", aquellas que son capaces de aceptar una gran diversidad de sustratos y que, por lo tanto, pueden ser usadas en varios procesos industriales. Centrándonos en éster-hidrolasas y ω-transaminasas de la clase III y aplicando y diseñando herramientas de cribado y tecnologías de alto rendimiento, hemos construido una de las mayores colecciones de estas enzimas localizadas en un solo laboratorio, que actualmente asciende a 155 (145 y 10, respectivamente). Gracias a su intensa caracterización con al menos 130 sustratos diferentes, se ha evaluado la especificidad y selectividad de sustrato, e identificado aquellas con características más prominentes. Con la ayuda de las estructuras y herramientas computacionales, se han identificado marcadores de "promiscuidad de sustrato" y enantio-selectividad, a partir de los cuales hemos sido capaces de acelerar la identificación de éster-hidrolasas y ω-transaminasas de la clase III versátiles e industrialmente relevantes, en la medida en que presentan especificidades de sustrato y selectividades mayores y poco frecuentes en comparación con enzimas comerciales similares. A continuación, se demostró que es posible transformar una enzima con amplia "promiscuidad de sustrato" pero no estéreo-específica en un biocatalizador estéreo-específico con amplia especificidad de sustrato aplicando técnicas de ingeniería supramolecular e ingeniería de proteínas. Los principales logros de esta Tesis incluyen: I) la generación de un set robusto de metadatos que demuestran que la biodiversidad ambiental es una valiosísima fuente de enzimas novedosas; II) la acumulación de datos que demuestran que es posible identificar enzimas con alta "promiscuidad de sustrato" y que, a su vez, son más comunes de lo esperado; III) la demostración de que el nivel de "promiscuidad de sustrato" de ciertos tipos de enzimas puede ser predicha a partir del análisis de sus secuencias, sus modelos tridimensionales y/o sus estructuras, sin necesidad de su clonación, expresión y caracterización previa; IV) reforzar la idea de que una amplia especificidad de sustrato va ligada a una baja enantio-selectividad, si bien este hecho puede solventarse mediante ingeniería supramolecular, restringiendo la flexibilidad de la estructura de la enzima mediante estrategias de inmovilización; V) la demostración de que es posible introducir un segundo centro activo a una enzima localizando previamente nuevos sitios de unión al sustrato, y que generar una enzima con dos centros activos funcionales podría mejorar la competitividad y las propiedades catalíticas de los biocatalizadores.

FIGURE INDEX

ABBREVIATIONS

3D: Three dimensional
Arg: Arginine
Asp: Aspartic acid
ATA: Aminotransferase
BSC: Barcelona Supercomputing Center
CSAR: Community structure-activity resource
CSIC: Consejo Superior de Investigaciones Científicas
EC: Enzyme commission
EH: Ester-hydrolase
FAO: Food and Agriculture Organization
GFP: Green fluorescent protein
GOS: Global ocean expedition
HAD: Haloacid dehalogenase
His: Histidine
ICP: Instituto de Catálisis y Petroleoquímica
Ile: Isoleucine
Leu: Leucine
MCP: Meta cleavage product
MOF: Metal-organic framework
NMR: Nuclear magnetic resonance
PCR: Polymerase chain reaction
PELE: Protein Energy Landscape Electron Exploration
Phe: Phenylalanine
PLP: Pyridoxal-5´-phosphate
PMP: Pyridoxamine-5´-phosphate
SASA: Solvent accessible surface area
Ser: Serine
Trp: Tryptophan
Tyr: Tyrosine
UNESCO: United Nations Educational, Scientific and Cultural Organization
WHO: World Health Organization
ω-TAs: ω-Transaminases

INTRODUCTION

1.1. Environmental and human health crisis

The design and use of tools have been key factors in human evolution and its relationship with the environment. By applying different tools, we can adapt or find solutions to environmental concerns, while at the same time provide improvements in a large number of different fields, such as agriculture or medicine. These tools and the developments generated from them are crucial given that we are about 7.7 billion living people, as most recent United Nations estimates. Even though it is true that the current growth rate is not as high as it was at other times in history, like at the industrial revolution, we will reach 10 billion by 2057 (https://www.worldometers.info). This overpopulation necessitates more technological advances in sectors such as Biotechnology, to give a specific relevant field to this Doctoral Thesis, to maximize the utilization of available ecological resources and sustain our way of life, given the estimates that the so-called Earth Overshoot Day is reached earlier year by year (https://www.overshootday.org/).

In addition to the overexploitation of limited available resources, overpopulation has a profound impact on our environment. Our increasing influence on the climate by burning fossil fuels, transportation, deforestation and chemical pollution, has made several authors name our current geological era the Anthropocene. Just as an example, the concentrations of typical greenhouse gases (carbon dioxide, methane or nitrous oxide, to cite some) are highly increasing because of human activities, being 40% higher than they were at the beginning of industrialization (Pachauri, 2014). These and other contaminants, such as oxygen-consuming materials, heavy metals, persistent organic matter and pesticides, are and will continue affecting air, soil and water reservoirs. Their presence will have a negative legacy of pollution across the planet and will affect human health on a global scale (World Health Organization (WHO), 2017). Just in 2019, exposure to air pollution was found to be associated with a circa 1% increased respiratory and cardiovascular mortality (Liu et al., 2019). One relevant example in relation to environmental concerns is plastic, widely used in our economy and of which 4-5 million tons accumulate annually in the oceans. Still, a biotechnological solution to recycling plastic is awaiting. In addition to this, the global average temperature is continuously increasing, being about 0.82°C higher than it was at the end of the 19th century. The consequences of the global warming include the rising of the sea level, the flooding of coastal areas, the dysregulation of rainfall, heatwaves and droughts and the occurrence of catastrophes, the increase of environmental pollution, and the negative impact on human economy and health.

Remarks: As we are responsible for the environment and resources crisis, we do need not only technological products but also a sustainable way to manufacture them as we are now dependent on them. This is essential in order not to plunder natural resources which are finite (https://ec.europa.eu/clima/change/), while at the same time minimizing the impact in environment and healthcare. These are among the reasons for the so-called latest *Biotechnology Boom*.

1.2. Biotechnology Boom

Biotechnology is the implementation of live organisms, or parts of them, in an industrial setting. Although the term was first coined by Karl Ereky in 1919 (Ereky, 1919), we had been using biotechnology from the firsts fermentations for making dairy products or alcoholic beverages (Bhatia & Goli, 2018). Nowadays it is presented as a sustainable alternative that can reduce costs in many industrial sectors. No one can deny the positive growth rate of this technology since a few decades ago a bunch of enzymes was implemented in the market (Arbige & Pitcher, 1989). As recently published, 2018 was the year of Biotechnology (Hodgson, 2019), according to the fact that private biotech companies raised more money in 2018 than

in any previous year (**Figure 1**). In addition to that, the global bio-economy, which depends heavily on biotechnological developments, is growing at a rate of about 3.6%, and this rate will continue for at least the next five years (Sawaya & Internacional , 2010; Sheldon & Brady, 2019).

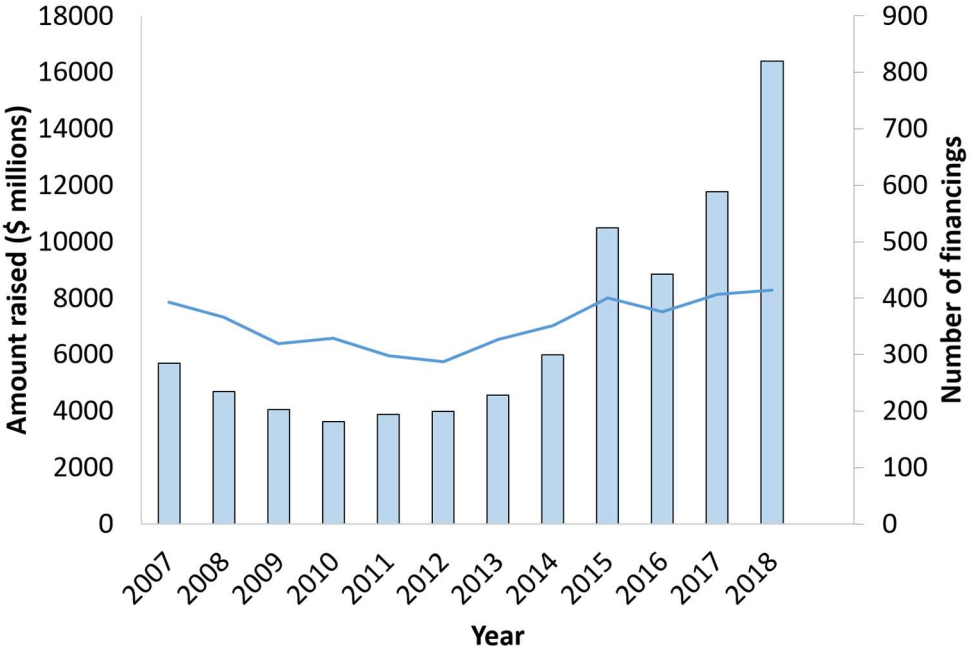


Figure 1. Global amount and number of financings raised by venture capital funds, 2007–2018, in Biotechnology. Modified from Hodgson (2019).

Within the field of Biotechnology, the emerging "White Biotechnology" has the mission to largely implement microbes and the enzymes they contain in place of conventional chemical synthesis (Bornscheuer et al., 2012). About 90% of all processes in industry require a catalytic step, and it is expected that up to 20-40% of chemical synthesis processes, which commonly rely on the use of environmentally damaging bulk organic solvents and energy-demanding high-pressure and high-temperature processes, could be substituted by microbial and enzymatic processes in coming years (Sheldon & Brady, 2019). This is why it is recognized that the future of our bio-economy will be based on the use of known or yet to discover microorganisms and their enzymes. Indeed, companies use microbes and their products (enzymes) because they can lead to lower production costs or because they reduce time to market, which is highly important for producing compounds under price pressure like fine chemicals and pharma products. In addition to that, these organic catalysts give, in many cases, high specificity/enantio-selectivity and catalytic efficiency (Blomberg et al., 2013; Mate et al., 2019), and can be used at ambient temperature and pressure (Turner & Truppo, 2013).

Remarks: We want to believe, or at least they have been telling us, that the future of our bio-economy may come from new microbes and enzymes (Pellis, Cantone, Ebert, & Gardossi, 2018). Nowadays, the technology is enjoying such a high level of sophistication that it has enabled the generation of a spectacular vision of the microbial and enzymatic world never achieved before. This, together with the multiple possibilities for better exploring and exploiting the diversity of microbes and their enzymes to meet the needs of the industrial growth, make the possibility of better exploring the environment to produce future bio-based chemicals closer. However, the problem to be solved is not how to access and sample our environment and microbes, but to better recover the right microbe, the right chemical and the right enzyme and get them into the market (Bommarius, 2015; Nature Catalysis, 2018).

This is one of the targets of this Doctoral Thesis, which has focused on the investigation of microbial enzymes, namely, ester-hydrolases and class III ω-transaminases, and how through an accumulated biocatalytic, structural and computational knowledge, one can shorten the way to find or engineer those. This may partially cover the needs that have taken us to the Biotechnology Boom.

1.3. Enzymes as biotech enhancers: ester-hydrolases and amine transaminases as study case

Enzymes are proteins that act as catalysts, and as such, they are reaction accelerators. In this case, this acceleration is due to a three-dimensional space, called active site, that within a polypeptide scaffold creates a biochemical ambient in which the reaction becomes energetically favorable. This is why chemical- and bio-industries have a steadily growing demand for enzymes that can catalyze a huge variety of different reactions and, if possible, with high activity, broad substrate specificity and stringent stereo-specificity. Using enzymes instead of chemical catalysts has some interesting advantages such us performing reactions at physiological conditions of temperature or pH, and avoiding or diminishing the release of damaging pollutants to the environment. However, since enzymes have been evolved by nature for working in living cells and under mild reaction conditions, most enzymes do not operate cost-efficiently in industrial processes conditions.

Despite the inherent limitations of enzymes, the global enzyme market is continuously increasing. Thus, the global enzyme market is expected to increase from 12.1 billion USD (in 2019) to 16.9 billion USD in 2024 (**Figure 2a**). This increase, which will occur in all major classes of enzymes currently on the market, including proteases, carbohydrases, lipases, polymerases and nucleases (**Figure 2b**), is directly linked to the necessity to produce high-demand products.

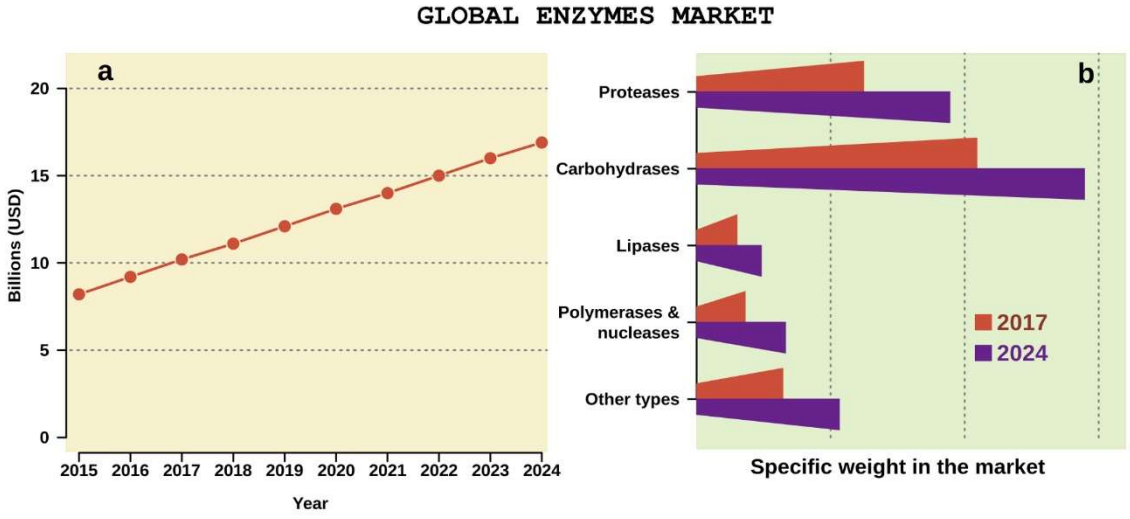


Figure 2. The market is increasingly demanding enzymes. a) Increase in the global market of enzymes in the industrial sector. b) Increase in the global market of each of the enzyme classes of relevance for industry (Market, 2016)

It is noteworthy that 64% of the World's enzyme-producing companies are located in Europe and that they produce around 75% of the world's industrial enzymes. Fifteen years ago, the global market of products obtained by biocatalysis and microbial fermentation in the sectors of food, pharma, fine chemicals, cosmetics, energy and new materials, had a global value of USD 70 billion. From now to 2023 it is estimated this value will increase five- to tenfold (Market, 2016).

This Thesis focuses on serine ester-hydrolases and class III ω-transaminases, two of the most attractive and used enzymes in industrial settings (Daiha, Angeli, de Oliveira, & Almeida, 2015; Ferrer et al., 2015). Below, an overview of their characteristics, mode of action and applications is briefly described.

Ester-hydrolases: what is known and what is missing

Ester-hydrolases, including esterases and lipases, represent a very important and, probably, one of the most widely used classes of enzymes for industrial biocatalytic applications, including kinetic resolution by enantio-selective hydrolysis and transesterifications, and, albeit currently less prominent, enzymatic polymerization (Ferrer et al., 2015). They are found interesting in food processing, beverages, laundry or perfume industries, because of its role in the processing of some esters with flavor and fragrant qualities and lipids. As well, they are widely used in pharmaceutical industries for the production of chiral drugs, in agriculture as they can react with the toxic residues of the phosphotriesters that are used as polluting insecticides, in paper industry to treat the paper pulp, in the textile dyeing process and production of leather, and in baking industry, among other relevant applications and processes (Panda & Gowrishankar, 2005; Daiha et al., 2015; Romano et al., 2015; Javed et al., 2018; Ozer, Ay Sal, Belduz, Kirci, & Canakci, 2019).

Esterases and lipases, with the Enzyme Commission numbers EC 3.1.1.1 and EC 3.1.1.3, respectively, are widely distributed through plants, animals and microorganisms. They are capable of breaking ester bonds and, in most cases, do so without any molecule as a cofactor. These properties make esterases and lipases an attractive option for industry (Sayali, Sadichha, & Surekha, 2013). From a structural and mechanistic point of view, these enzymes have a typical α/β hydrolase fold with, in most cases, a catalytic triad formed by Ser-His-Asp, assisted by the so-called oxyanion hole, in the particular case of serine ester-hydrolases (**Figure 3**).

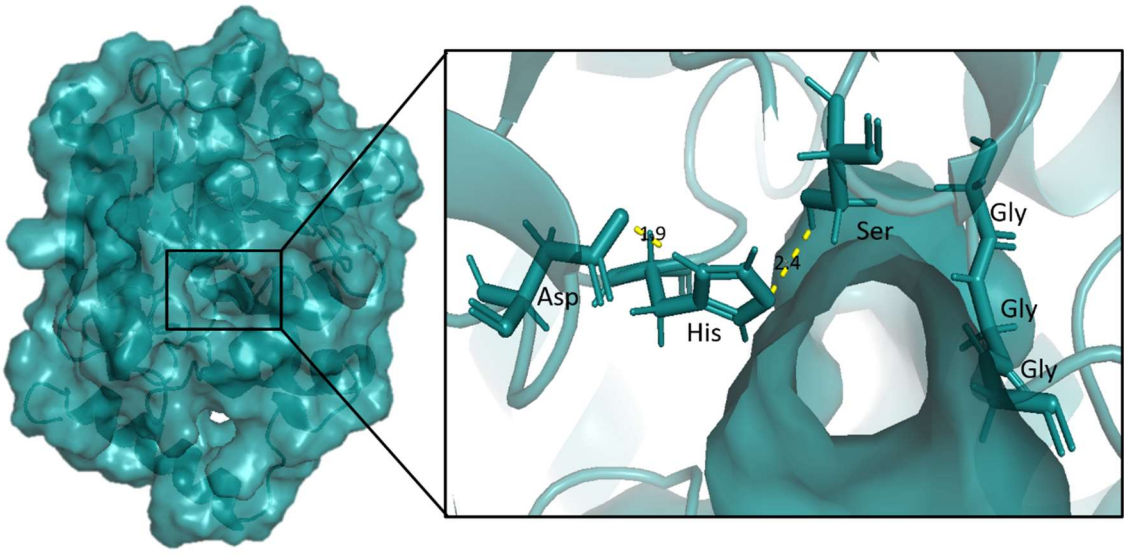


Figure 3. Detail of the Ser-Asp-His catalytic triad and Gly-Gly-Gly oxyanion hole, in the free EH1 serine ester-hydrolase (PDB 5JD4) reported in this Thesis. Distances between triad residues are shown in Angstroms.

The currently accepted mechanism of the ester rupture by serine ester-hydrolases needs one molecule of water. In resting conditions, the Serine has a nucleophile oxygen due to the Histidine acceptance of a proton. The Histidine can accommodate this extra positive charge as the Aspartic acid sidechain is negatively charged in neutral pH and relies very close to the Histidine. This nucleophile oxygen of the

Serine can be attached to the carbonyl carbon of the ester by releasing an alcohol molecule. An acyl tetrahedral intermediate compound is formed and attacked by a water molecule in a second nucleophilic reaction, breaking the union between enzyme and substrate and releasing the acidic group. The stabilization of the interaction is mediated by the so-called oxyanion hole contacts, commonly constituted by adjacent Glycine residues with nitrogen atoms. A detailed scheme of the interactions and reaction mechanism is shown in **Figure 4**. For extensive details see a recent review by Aranda et al. (2014).

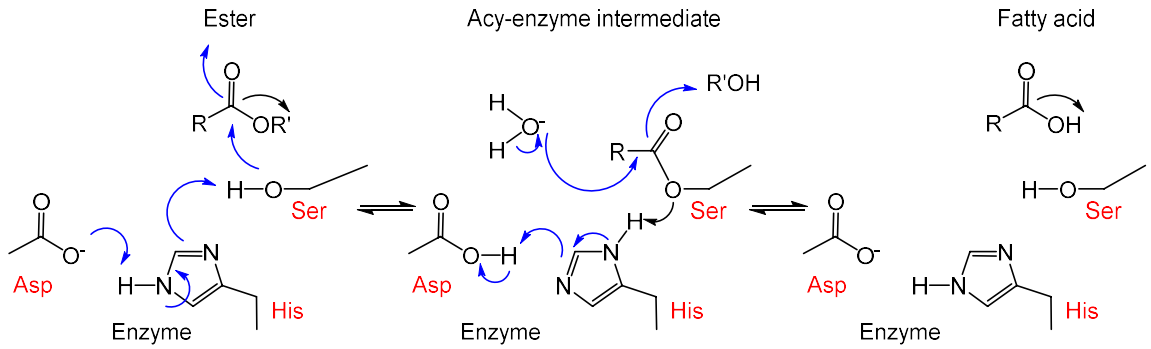


Figure 4. Scheme of the model catalytic mechanism of ester hydrolysis by serine ester-hydrolases. Adapted from Aranda et al. (2014).

Based on the comparison of primary structures, at least 14 different families of esterases and lipases have been described in databases, each containing highly diverse enzymes differing at sequence, structural and properties levels (Ferrer et al., 2015). They are widely distributed in the environment and within the phylogenetic tree and are relatively easy to screen by multiple methods (Adesioye et al., 2018) and to clone, produce and characterize (see Chapter 1) compared to other enzyme classes. In addition to that, random mutagenesis (Adesioye et al., 2018), rational protein engineering (Hajighasemi et al., 2016) and immobilization (Keller, Beloqui, Martínez-Martínez, Ferrer, & Delaittre, 2017) had served to improve stability, enantio-selectivity and process performance. However, these strategies, albeit a number of successfull stories have been reported, appear to be limited to a small group of specialists rather than having the potential for everyday application in industry. Regardless of this situation, a great amount of knowledge of their sequences and biocatalytic properties has been accumulated, thanks to which multiple preparations are commercially available for industrial purposes. For extensive details see a recent review by Ferrer et al. (2015).

Although several esterases and lipases for biotechnological applications have already been identified, being CalB from *Pseudozyma aphidis* (formerly *Candida antarctica*) one of the best representatives of commercially available lipase preparations used for the production of fine chemicals (Daiha et al., 2015), the molecular basis of their performance remains largely unknown. Thus, such enzymes, as all others, are usually identified using surrogate substrates, and once confirmed their activity with such model chemicals, their capacity to convert other molecules is further confirmed experimentally. Unfortunately, the ability of an esterase/lipase to catalyze a pre-defined reaction is currently not predictable. Consequently, the first step towards a novel application is a laborious and time-consuming screening of large enzyme libraries for the desired activity (Ferrer et al., 2016). In addition to that, robust enzymes, e.g. for fast production of molecules requiring high-temperature polyurethane networks (e.g. at 80°C within 60 minutes) starting from oligomeric building blocks, are missing.

For all the reasons above, ester-hydrolases were used as the first objective along this Doctoral Thesis. We think that a major bottleneck in relation to ester-hydrolases is not their screening or the limited number of available preparations but rather the establishment of the molecular mecanisms responsible for catalytic tuning and how this understanding can help finding and defining the right one among many others with limited application potential. The idea is to start with a large screen program of such enzymes

with different proven substrate promiscuity and use this as a lead scaffold to first, understanding how substrate specificity is modulated and to secondly, engineering stability and selectivity towards the desired reaction. This is achieved by a synergistic combination of state-of-the-art structural, modeling, computational and protein engineering and immobilization. Gene sequences encoding enzymes to start with have been identified from deep mining metagenomes and bacterial genomes. A key step of this Thesis comprises ester-hydrolases identification and substrate selectivity evaluation to enable a rapid identification of novel and robust enzymes, establishing a pipeline trained with the top candidate enzymes. For that, the respective genes have been expressed using a unique collection of different host strains and plasmids to validate the predicted activities and to purify sufficient amounts for activity tests and crystallization studies. Solved crystal structures or 3-dimensional models were used to enable understanding of enzyme features as well as to engineering specificity and selectivity in benchmarking reactions.

Remarks: A high number of ester-hydrolases has been described, some of which are commercially available and used. Their screening is no longer a problem, because the existence of easy screen tests (Reyes-Duarte, Coscolín, Martínez-Martínez, Ferrer, & García-Arellano, 2018), and because they are relatively easy to produce. The bottleneck is how to find the ones with more application potential amongst many others. This Thesis will offer a highly diverse set of ester-hydrolytic enzymes, functional high throughput screening platforms to screen for substrate promiscuity, detailed knowledge of factors determining substrate promiscuity, and top candidate enzymes showing substrate promiscuity and activity that can constitute the starting point to build up a modeling/docking/crystallization pipeline to deepen into structural-functional analyses.

Transaminases: what is known and what is missing

Transaminases (EC 2.6.1.x), also called aminotransferases (ATA), are the second most requested group of enzymes at industrial scale for chemical synthesis (Ferrer et al., 2015; Ferrer et al., 2016). Relevant examples through pharmaceutical industry include the use of ATA-117 D-amino acid aminotransferase for producing Januvia™ (Sitagliptin phosphate), a drug for lowering sugar level in blood (Bommarius, Blum, & Abrahamson, 2011), and the use of some transaminases for the amination of carbohydrates for biomaterials which are typically made from fossil fuels (Aumala et al., 2019). They are versatile enzymes with industrial potential to produce building blocks for drug discovery and chemical biology, as they catalyze asymmetric amino? transfer reactions between an amine and a ketone, aldehyde or keto-acid. They use pyridoxal-5´-phosphate (PLP) as a cofactor, which serves as a molecular shuttle for ammonia and electrons between the amine donor and the acceptor (Espaillat et al., 2014; Fuchs, Farnberger, & Kroutil, 2015; Steffen-Munsberg et al., 2015).

Transaminases are classified into six classes (I to VI) based on their structural features and sequence identity, with class III, covering the so-called ω-transaminases (ω-TAs), being one of the most used classes in biotechnology, as they produce pure chiral amines that serve as building blocks in drug development (Steffen-Munsberg et al., 2015). PLP covalently binds to a Lysine via a Schiff base forming an internal aldimine. The catalytic mechanism is made up of two half-reactions. In the first one, the donor substrate forms a Schiff base with the cofactor and after further intermediate steps, transfers its amino group to the co-factor forming an external aldimine which generates pyridoxamine-5´-phosphate (PMP) when deaminated product is released. In the second half-reaction, the amino group is transferred from PMP to an acceptor ketone or aldehyde restoring the PLP internal aldimine. A scheme of the previously described transamination process is shown in **Figure 5**. For extensive details, see recent reviews by Sayer et al. (2014) and Steffen-Munsberg et al. (2015).

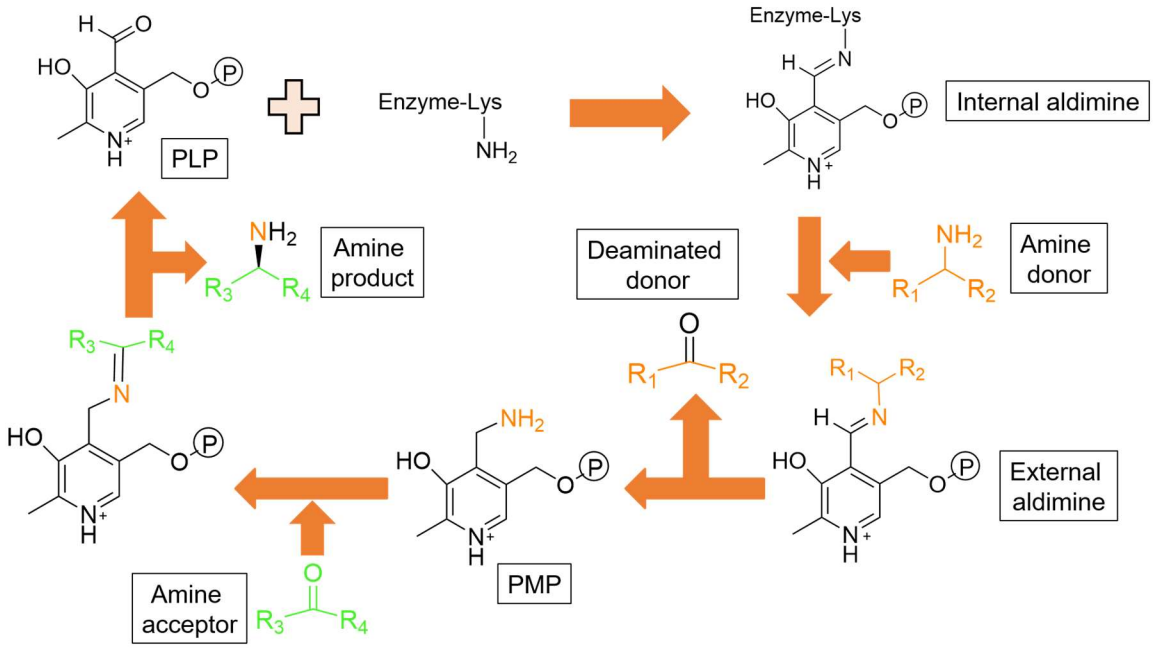


Figure 5. Scheme of global transamination reaction by class III ω-TAs. Image adapted from Sayer et al. (2014).

Although a Lysine residue is the primary catalytic amino acid, few other residues have been shown to be conserved among class III ω-TAs and are located in the catalytic pocket of the protein, commonly being a dimer (**Figure 6**). Aside Lys, essential to mediate the transaminase reaction as a catalytic base, these residues included: I) Tyr, Phe, Tyr and Arg residues which appear to have a sterically limiting role in the so-called L pocket of the active site and in the recognition of those substrates containing carboxylates (particularly, the Arg); and II) Leu, Trp, and Ile residues which create the steric constraints in the S pocket (Park, Kim, & Shin, 2012; Han, Park, Dong, & Shin, 2015). Substitutions in any of these residues have been shown to have a role in the substrate spectra, through generating additional size rooms required for bulky substituents. A schematic representation of the location of the active site in a model class III ω-TAs with an overview of residues most likely implicated in catalysis and substrate recognition is shown in **Figure 6**.

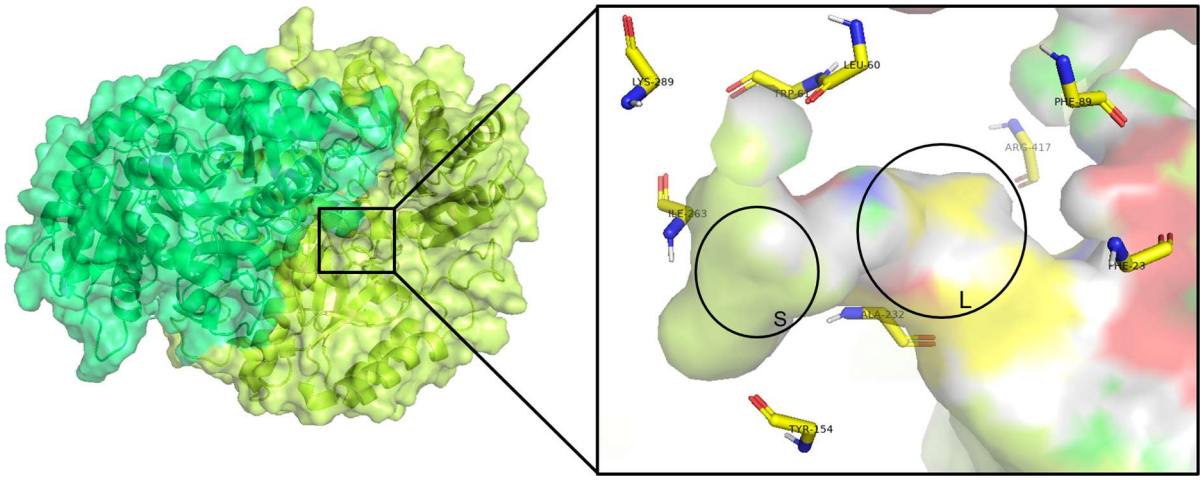


Figure 6. Structure of a class III ω-TAs called TR2 (see Chapter 3), and a detail of the catalytic active site of one of the two monomers with conserved amino acids that are forming the L-pocket and the S-pocket.

Compared to ester-hydrolases, two major bottlenecks are worth mentioning in transaminases research. First, amine transaminases from the class III ω-TAs are key enzymes for modification of chemical building blocks, but finding those capable of converting bulky substrates, including bulky amines and bulky

ketones, and (R)-amines is still challenging. Second, their discovery is limited most likely due to a lack of suitable screening methods, both experimental and *in silico* screens, at large scale.

Remarks: Although several aminotransferases with interesting properties and for biotechnological applications of relevance have already been identified, a major bottleneck is the discovery of novel such enzymes capable of converting bulky ketones and aldehydes and (*R*)-amines; establishing the molecular basis of such substrate preference remains largely unknown. For all these reasons, amine transaminases of the class III ω-TAs were a second objective along this Doctoral Thesis. The identification of new such enzymes and the extensive analysis of their substrate spectra may help to deepen into the molecular basis determining what makes such enzymes not commonly capable of converting bulky substrates and (R)-amines.

1.4. Finding or engineering the right enzyme: some of the desirable properties

Often, enzymes are more delicate or unstable than chemical catalysts, as they need to maintain their native structure to perform certain reactions that require incompatible pH or temperature conditions or the addition of solvents. Moreover, enzymes need to be commonly mutated to increase the yield or reaction rate to become industrially useful (Arnold, 2019). Enzymes normally take a lot of time to enter the market since we have a big amount of candidates, but in the end just a few of them are suitable for our purposes. In order to obtain an optimized bio-catalyst for a desired reaction, one can design biocatalysts *de novo* (Richter et al., 2012), improve our existing biocatalysts through enzyme engineering (Höhne, Schätzle, Jochens, Robins, & Bornscheuer, 2010; Arnold, 2019) or explore the current biodiversity in search of enzymes that are already performing this reaction under the conditions we are looking for (Ferrer et al., 2016). It is difficult for enzymes to enter the market but not impossible. As an example, the incredible advance in molecular biology thanks to Polymerase Chain Reaction or PCR technique was linked to the usage of the DNA polymerases from thermal resistant microorganisms as *Thermus aquaticus* which lives in hot springs. Also, the green fluorescence protein GFP widely used as a reporter was extracted from the jellyfish *Aequorea Victoria*. It is also to note the case of the lipase Lipozyme® CALB, initially commercialized by Novozymes A/S.

Enzymes with outstanding properties that may contribute to solving some of the major challenges of our time are difficult to find and predict. Promiscuity is one of these properties, which can be defined in different ways (**Figure 7**) depending on the characteristic we pay attention to (Copley, 2003). Among the different promiscuous phenotypes, this Thesis related to enzymes with the possibility of accommodating different molecules in the active site, performing the same overall reaction; this phenotype is referred to as “substrate promiscuity”.

The traditional statement that one enzyme can recognize only one substrate is not that solid anymore, as enzymes with broad substrate specificity (herein referred to as “substrate promiscuous” enzymes) have become more and more common (Nobeli, Favia, & Thornton, 2009). This phenomenon is highly relevant for microbes as their presence is linked to environmental (Huang et al., 2015), evolutionary (Huang et al., 2012; Yip & Matsumura, 2013; Price & Wilson, 2014; Devamani et al., 2016), and structural (Hult & Berglund, 2007; Copley, 2015; London et al., 2015) adaptations. However, in the context of this Thesis, this property is crucial from a biotechnological point of view as a narrow substrate spectrum is one of the most frequent problems for industrial enzyme applications (Nobeli et al., 2009; Ferrer et al., 2016).

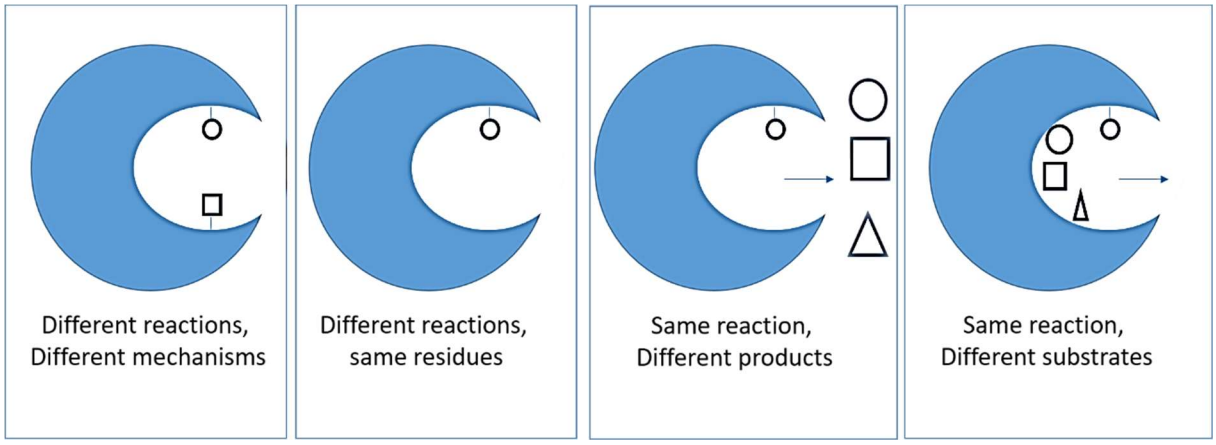


Figure 7. Classification of "promiscuous" phenotypes of enzymes.

Remarks: A narrow substrate spectrum is one of the most frequent problems for industrial enzyme applications (Ferrer et al., 2016). A consensus exists that "the more substrates an enzyme converts the better", opening application ranges with the consequent reduction of the production cost of multiple enzymes.

Across this Doctoral Thesis, we first focused on the investigation of substrate promiscuity as a relevant enzymatic property to work with and understand. There is a number of current challenges in relation to understanding substrate promiscuity. Enzymes with wide substrate ranges appear naturally (see Chapter 1 and references therein). Indeed, some enzymes are more promiscuous than others for multiple reasons. In fact, it is generally accepted that the substrate specificity of individual enzymes is influenced by the architecture (size and geometry) of their active-site cavity and their access tunnels (Damborský & Koca, 1999; Suplatov & Švedas, 2015). The elaboration of a core domain or minimal structural unit, from which these properties can diverge, may have led to an increased substrate range in a superfamily (Huang et al., 2015; Berezovsky, Guarnera, Zheng, Eisenhaber, & Eisenhaber, 2017). In general, large active sites are consistent with these enzymes' very broad substrate specificity, whereas the enzymes with smaller and occluded cavities cannot accommodate so readily larger substrates (Suplatov & Švedas, 2015). Plasticity is also an important factor, as enzymes with different degree of plasticity showed altered substrate conversion, since the active site may acquire diverse conformations (Albesa-Jové & Guerin, 2016). The presence of key substitutions in the access tunnels may also control the specificity, as it has been demonstrated by protein engineering and comparative computational studies (Espaillat et al., 2014). These substitutions and their positioning may increase or decrease the activity towards certain substrates not only because they influence the entrance and positioning of the substrates themselves (i.e., steric hindrance), but also by restricting the access of water molecules (Koudelakova et al., 2011) or contributing to the positioning of different anions in the proximity of the active site (Albesa-Jové & Guerin, 2016; Purg et al., 2016). This may alter, for example, the rate of the hydrolysis of an acyl-enzyme intermediate formed during substrate hydrolysis by ester-hydrolases. Divergence in specificity may also be due not to plasticity and simple amino acid residues substitutions (Alcaide et al., 2013), but to large structural elements that are inserted, removed or rearranged in the primary sequence (Huang et al., 2015; Ramírez-Escudero et al., 2016). By comparing wild-type and mutant variants it has also been demonstrated that it is possible to interconvert the substrate specificity of enzymes by modifying their active-site cavities and main access tunnels (Alcaide et al., 2013).

Moreover, vast protein sequence comparisons and experimental data sets revealed the existence of enzymes with wide substrate specificity in many protein families (Kmunícek et al., 2005; Yoshida & Komae, 2006; Simpson et al., 2009; Gajda, Pawełczak, & Drag, 2012; Li & Shao, 2014; Huang et al., 2015; Wilding,

Peat, Newman, & Scott, 2016; Yang, Lin, & Wei, 2016), including lipases (Daiha et al., 2015). A systematic analysis of substrate specificities additionally revealed that phylogenetic analysis cannot be used to predict the substrate specificity (Koudelakova et al., 2011). Thus, case-by-case investigations are commonly needed to properly link a sequence not only to a function but also to a range of molecules capable of being accepted as potential substrates. That is, the level of predictability by which enzymes bind and convert multiple substrates is unknown, although molecular insights have been reported for single enzymes (Babtie, Tokuriki, & Hollfelder, 2010). How can one distinguish substrate promiscuous vs non-promiscuous enzymes, if possible? How can one precisely select an enzyme with a broad substrate range over thousands or millions of others checking its sequence and avoiding extensive cloning and activity tests? These are important initial questions to answer in order to find marketable enzymes. A tool that can clearly distinguish promiscuous vs non-promiscuous enzymes and suggest substrates potentially being converted or not by them might therefore be valuable to apply low-cost sequencing in discovery platforms in large scale datasets such as those produced by genomics and metagenomics, while at the same time leading to new versatile biocatalysts capable, at least, of converting multiple chemical scaffolds. This is particularly relevant when confronted with an enzyme encoded by a sequence one cannot *a priori* predict whether it will be capable of converting more or fewer substrates without previously characterizing it. This is the first pre-requisite for its potential commercial exploitation.

As mentioned above, the relevance of predicting substrate spectra is indisputable, but there are other properties as relevant for manufacturing settings as promiscuity. On the basis of specialized literature, and multiple collaborations with private companies, what Industries also need are "easy-to-produce enzymes displaying a high turnover rate for industrial model-molecules, the more, the better, with stringent regio- and stereoselectivity and specificity when confronted with molecules having multiple chiral or isomeric centers, and good stability". Therefore, at least, substrate promiscuity, substrate conversion rate, substrate specificity and selectivity and stability are fundamental properties to evaluate, understand, predict and improve on biocatalytic manufacturing settings. Among them, stereo-specificity is the second property investigated in this Thesis. Certainly, as described above, some enzymes are more promiscuous than others but can also convert substrates that others cannot or even convert enantiomers that others cannot. Are broad substrate promiscuity and stereo-specificity opposing properties in an enzyme? Can we find enzymes with broad substrate range but at the same time having stringent regio- and stereo-specificity? And, most importantly, can we answer these questions by examining sequence information without requiring their time- and cost-consuming cloning, expression and characterization? We think that these questions also remain unanswered and constitute a few crucial challenges coming along with the implementation of enzymes in industrial biotechnology.

In order for enzymes to become attractive for industrial processes development, it is crucial not only to identify those with properties of interest as mentioned above but also to improve them, as many of them have originally showed frustrating poor performance or are not stereo-specific under industrial conditions or requirements of productivity. Extensive genetic (Arnold, 2019) and supramolecular (Sheldon & Brady, 2019) engineering tools (empirical and *in silico*) are being applied and designed in this direction as the only way to definitively evaluate the effect of small structural or amino acid changes or flexibility constraints for the control and expansion of substrate binding, selectivity, and specificity (**Figure 8**). Although major advances in the field of engineering (genetic or supramolecular) and rational design are being achieved, a challenging question still remains in the topic of designing marketable enzymes: the absence of tools by which substrate spectra, selectivity, and specificity and enzyme stability could be inferred from sequence data which limits further enzyme improvement by rational design. If one has a tool from which substrate spectra, selectivity and specificity could be inferred, then it would be possible to direct the rational design of enzymes having appropriate requirements for Industry. However, this

possibility has not been developed. We think that identifying the best enzymes and designing tools for predicting enzyme properties could generate useful information to better guide protein design.

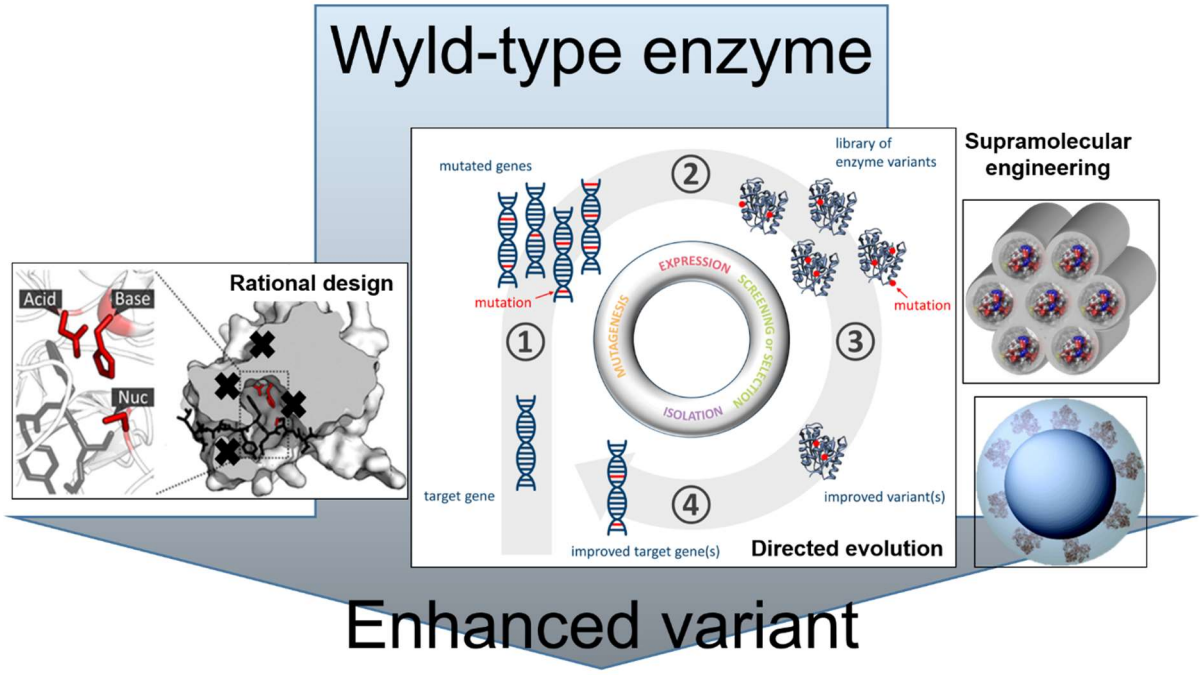


Figure 8. Some techniques that can be applied to improve the performance of wild type enzymes.

Remarks: Chemists mostly use inorganic catalysts for chemical conversions, and they commonly cannot imagine using enzymes as a first choice. Only if chemistry does not find a solution or there is an environmental concern around the chemical process, enzymatic conversions are considered at industrial level. How to find and predict enzymes entering to the market faster? In an ideal scenario, functional characterization of a large number of enzymes with genomic and metagenomic techniques using a large library of substrates, including industrial model-molecules, would guide the accumulation of biocatalytic data. These data may provide valuable information to comprehensively understand the mechanistic basis of promiscuity, selectivity and specificity and to precisely predict and direct the discovery of new attractive biocatalysts with broad substrate profile and stringent stereo-specificity. Besides this, it is generally assumed that natural enzymes cannot be directly used at industrial conditions and that one or several rounds of engineering or immobilization will be needed. This is why, it is not only important to find enzymes with relevant properties and how to predict their incidence, but also to apply and design new engineering and immobilization tools to tune their properties. Nowadays, one cannot anticipate which of the alternatives, engineering (at genetic or supramolecular level) existing enzymes, or finding them *de novo*, is a better alternative as a source of better performing biocatalysts, and this is why both approaches should be considered in any research and apply setting, as it was the case of this Thesis.

1.5. Source of enzymes: metagenomics as a tool to exploit biodiversity

Modern white biotechnology is based on the application of enzymes predominantly derived from microorganisms. Microbes dominate our planet considering their cell number, number of species, total biomass and diversity of environmental conditions in which they can grow (Mora, Tittensor, Adl, Simpson, & Worm, 2011; Kyrpides et al., 2014). It is therefore anticipated that yet unexplored microorganisms and their genomic and metabolic diversity (especially that from the extremophiles) will provide the material

to enable the most valuable new biotech applications (Berenguer & Ventosa, 2018). However, the classical biochemistry and biotechnology relied on the availability and culturability of single microbial isolates (Yarza et al., 2014). Although efforts in cultivation-based strategies to find new microbes relevant as biocatalysts are being developed (Rodrigues, Pereira, Fernandes, Cabral, & de Carvalho, 2017), the rate of discovery of biotech-relevant products is severely limited because we were unable to culture the vast majority of microbial species. Only a fraction of less than 5% are currently maintained in labs and culture collections; indeed, representatives of about 11000 species have been cultured (Yarza et al., 2010), and if we were going to isolate all microbial species predicted thus far, about 10 millions (Mora et al., 2011), it would take at least 1000 years.

This is why the existing and recognized potential for environmental microbiology to substantially improve the commercial potential of biotechnology has recently been greatly further strengthened by the advent of the "metagenomics" (Singh, 2010; Wilson & Piel, 2013; Ferrer et al., 2016; Ye, Siddle, Park, & Sabeti, 2019). Metagenomics was first described by Handelsman (Handelsman, Rondon, Brady, Clardy, & Goodman, 1998) as a group of genomic techniques applied to the microbiology biodiversity of an environmental complex sample skipping the laboratory culture step. Just around 5% of microorganisms can be cultured in laboratories so theoretically we were losing a great amount of biodiversity by using traditional techniques when searching for new enzymes (Puspita, Kamagata, Tanaka, Asano, & Nakatsu, 2012). **Figure 9** summarizes all the steps and techniques which, constituting the so-called metagenomic tools, can be applied to screen microbial communities for discovering new enzymes.

Metagenomics provides the capacity to discover new proteins in microorganisms and their communities without the need to culture them as individual species, which is technically very difficult and time-consuming (Puspita et al., 2012). Just as an example, the total time and cost of all the necessary steps for cultivating a new microbe varied from days to years time, and from few to dozen thousands euro, at least; easy-to-isolate aerobic bacteria requires at least one to two months to obtain a pure culture, with a cost for preparation of cultivation media, maintenance and identification of about 2,500 € per strain; culturable anaerobic strains would need 2-3 months, with a cost of 4,000-5,000 €; and difficult-to-isolate (micro)aerobic strains such as *Zetaproteobacteria* or *Thaumarchaea* would take 1-2 years and cost ca. 10,000 € each. Instead, with metagenomics, the DNA of microbial communities is directly harvested from the environment, sequenced and bioinformatic analysis and/or expression is then performed in surrogate microbial hosts enabling screening for enzymatic activities of interest, in days to weeks time. This has already been shown to facilitate the discovery of new enzymes with novel activities of industrial relevance (Ferrer et al., 2016).

Whether or not to perform pre-enrichments of the environmental sample before DNA isolation, use functional or *in silico* screens, use different vectors and hosts while cloning environmental DNA, and use different sequencing platforms, is at the disposal of researchers and different options are available. In addition, recent achievements in DNA sequencing and its price reduction, molecular biology and bioinformatics allow easy access to an enormous amount of enzymes. These techniques, their advances and the future of metagenomic data analysis in multiple contexts have been extensively described in literature (Popovic et al., 2015; Ferrer et al., 2016; Garrido-Cardenas & Manzano-Agugliaro, 2017; Ngara & Zhang, 2018; Ye, Siddle, Park, & Sabeti, 2019; Valdes, Stebliankin, & Narasimhan, 2019), and will not be detailed in this Thesis.

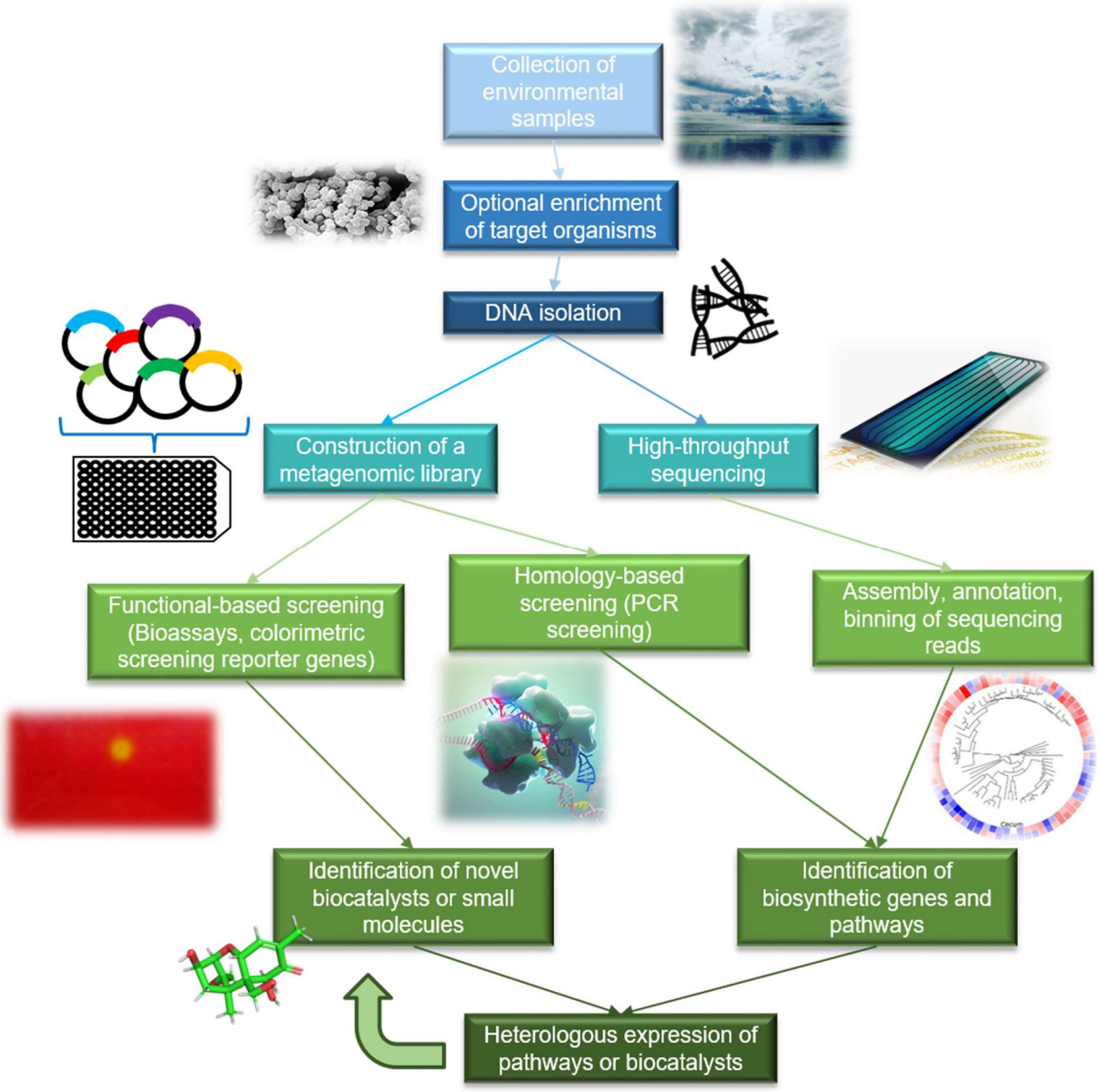


Figure 9. Metagenomic-based pipeline for enzyme discovery. Figure adapted from Wilson & Piel (2013).

Nowadays, enzymes can be efficiently identified and screened from metagenomic libraries or through homology searches in DNA databases (Ufarté, Laville, Duquesne, & Potocki-Veronese, 2015; Ferrer et al., 2016; Kunath, Bremges, Weimann, McHardy, & Pope, 2017; Montella et al., 2017). A set of a few hundred new enzymes can be found using a simple substrate within few months. The cultivation of metagenomic libraries under selective conditions allowing the growth of the wanted clones accelerates the finding of new enzymes significantly (Popovic et al., 2015). Hence, establishing an enzyme collection is no longer a big hurdle. In addition to that, genomes of cultivable microbes or metagenomes can often be easily inspected for such enzyme genes that are cloned and biochemically characterized. Eventually, the enzymes can be expressed at a pilot scale, validated and implemented into a process (Popovic et al., 2015). However, this last step is a very time consuming and expensive process that only a small number of enzymes will finally pass. A timeline of several years from the identification of a gene to process establishment is the reality rather than the exception (Ferrer et al., 2016). Based on real examples, at least 58-66 weeks would be required to discover and commercialize an enzyme from a microbe, from a starting number of several thousands of pre-selected enzymes. The total cost of the whole process, including

investigating whether a continuous supply of a novel enzyme at the right price is feasible, the discovery of the right enzyme, the production and delivery of an active enzyme with the desired technical specifications, and the production of a standard enzyme product with validated, regulated and quality-assured routine manufacturing may be at least 100,000 € each, with only 1 out of 100,000 recovering the cost (https://blog.oup.com/2019/04/industry-marine-diversity/).

Although the discovery of genes encoding for enzymes by metagenomic analysis is a first option because its relatively easy feasibility to screen enzymes, other -omic techniques such as metaproteomics are available for such purpose (**Figure 10**). As an example, the presence of enzymes can be checked by directly assessing activities in proteins extracted from environmental samples or enrichments and further, using mass spectrometry-based metaproteomics or metatranscriptomics by deep sequencing (RNAseq). By using these techniques, one can mainly identify enzymes with high turn-over rates on the substrates of interest (Sukul et al., 2017), information which cannot be obtained by DNA sequencing of the metagenome.

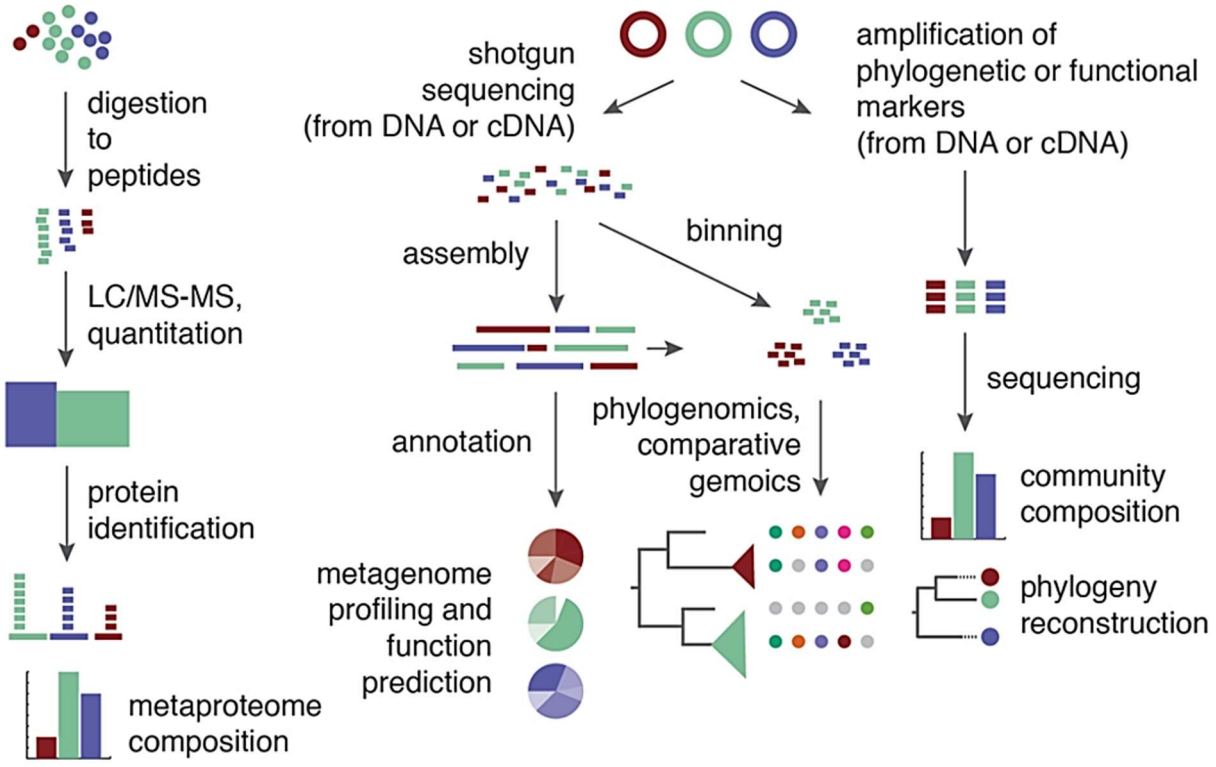


Figure 10. Simple discovery of bacterial biocatalysts from environmental samples through functional metaproteomics compared to metagenomic analysis.

Remarks: As the number of sequences deposited in databases is rising exponentially, there is an important gap between the number of sequences available and the number of enzymes characterized. For this reason, this Thesis is contributing to enzyme characterization efforts. This was done by incrementing the number of enzymes being isolated by metagenomic tools and by applying extensive characterization and computational tools, selecting those with most interesting properties and searching for structural patterns that tune substrate specificity, which was further improved through rational design and supramolecular modification of the scaffold flexibility through immobilization.

1.6. Study of structural patterns for predictive markers and design of better enzymes

As mentioned above, the relevance of predicting promiscuity level and/or substrate spectra is indisputable. As discussed in Section 1.4, case-by-case investigations are commonly needed to properly

link a sequence, not only to a function but also to a range of molecules, isomers and chiral molecules capable of being accepted as potential substrates. A similar situation is encountered when evaluating and trying to predict solvent stability, as no available tool allows *a priori* prediction of solvent stability from sequence data. Despite this, the temperature stability of an ester-hydrolase from *Bacillus subtilis* was recently effectively predicted with computational methods applying rigidity theory-based Constraint Network Analysis. However, this method can only be used with small proteins/short peptides and whether this method can be extended to other ester-hydrolases or enzymes needs to be evaluated (Rathi, Fulton, Jaeger, & Gohlke, 2016). With the goal of making better enzymes according to industry application requirement of high stability and high activity, other automated strategies that design stable networks of interacting residues at the active site have been applied for improving activity and stability of hydrolases (Khersonsky et al., 2018).

Remarks: An accumulated knowledge of enzyme properties will allow a better predictability of the best enzymes, how to easily find them in a database, and how to use this information to produce and design a set of commercially exploitable enzymes. Recent advances in computational tools, including modeling and docking can contribute to the analysis of meta-biocatalytic data.

Computational methods to find predictive markers: from modeling to docking

In order to approach the analysis of meta-biocatalytic data, finding predictive molecular or structural markers responsible for an enzyme property and the design of predictive tools around them, a number of data management, bioinformatic and computational tools needs to be commonly considered. First, bioinformatic or data management tools need to be in place to have a comprehensive understanding of the general molecular bases underlying what determines each of the enzyme properties. For example, hierarchical clustering and network analysis of enzymes properties to evaluate the differences in properties associated with each enzyme are needed. The clustering, which can be fed from the metadata, can be created with the R language console, and the hierarchical clusters generated by calculating a distance matrix using a "binomial" method, the hclust function, and the functions as.phylo and plot.phylo from the ape package. Also, a sequence-to-properties network analysis can be performed by the 'organic' layout algorithm in Cytoscape (version 3.2.1), where clusters or subgraphs, each representing different relationships, can be identified. A number of parameters characterizing the primary and tertiary protein structures can be calculated, and further included in the hierarchical clustering and network analysis. While some of these parameters can be calculated from the sequences, such as sequence identity, family assignations, amino acid composition and isoelectric point (pI), others need three dimensional (3D) model or structural analysis.

One of the reasons for price reduction of sequencing was the development of bioinformatic tools with computing power enough to manage and store the great amount of data and its accessibility in public databases (Diniz & Canduri, 2017). As bioinformatics evolve, new software programs with new functions appear. In order to perform structural analyses, it is imperative to know the protein structure of our enzyme. However, protein structures are very difficult and costly to obtain and often three dimensional (3D) structural patterns of enzymes with unknown structures are studied using homology modeling software programs such as Swiss-model. The 3D modeling of an amino acid sequence starts from the alignment of it against the protein sequence of an already characterized structure, as schematically shown in **Figure 11**.

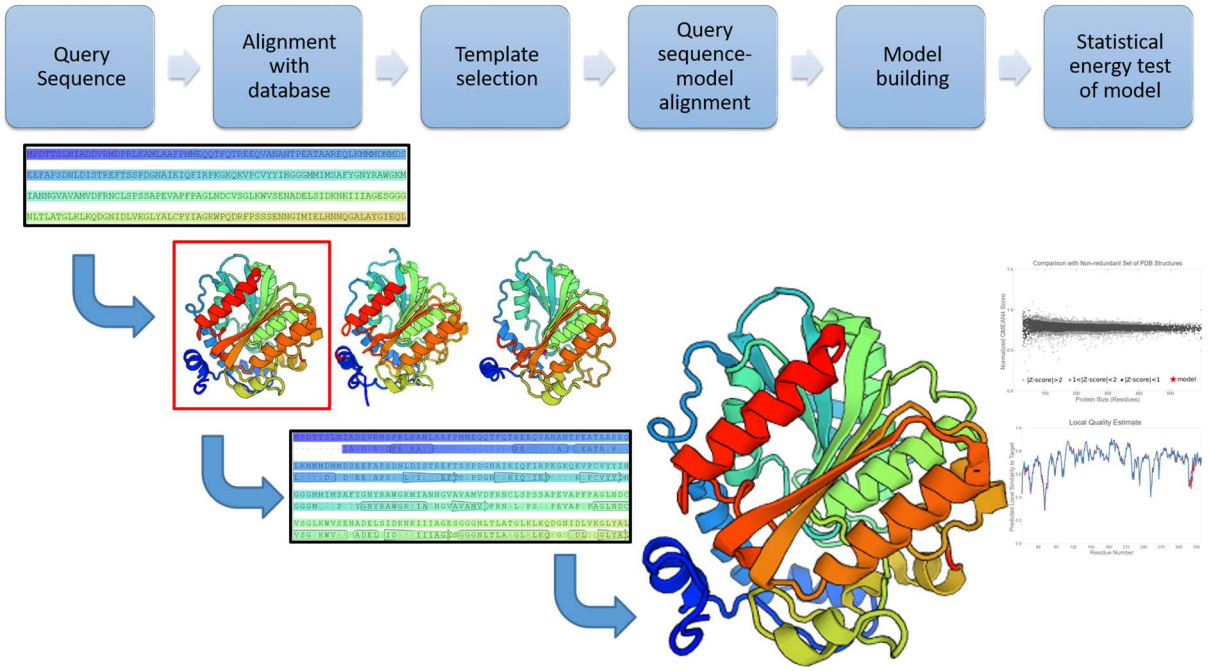


Figure 11. Schematic representation of the 3D structural analysis using the 3D modeling software Swiss-model (https://swissmodel.expasy.org/).

Through this Doctoral Thesis, we have used not only crystals but also models of our proteins for studying the structural patterns in the context of biocatalytic properties, namely, substrate promiscuity. In addition, for topology studying catalytic pockets, we have used the Protein Energy Landscape Exploration (PELE) software (Borrelli, Vitalis, Alcantara, & Guallar, 2005) for docking probes. PELE offers today one of the best modeling alternatives to map protein-ligand dynamics and induced fit. This technique combines protein structure prediction methods with a Monte Carlo algorithm, and represents a breakthrough in modeling protein-ligand interactions, being recently highlighted in the latest community structure-activity resource or CSAR challenge (a blind benchmark for docking and scoring methods) as a "remarkable achievement" (Borrelli, Cossins, & Guallar, 2010; Hernández-Ortega et al., 2011; Carlson et al., 2016; Santiago et al., 2016). Furthermore, the new Adaptive-PELE (Lecina, Gilabert, & Guallar, 2017) allows for a complete protein surface exploration, locating binding site pockets in only a few hours of a moderate computing cluster (~32 computing cores). Thus, these technological developments are ideally suited to locate, without previous knowledge, potential binding sites that could be converted into active sites. We should underline that finding these pockets typically requires significant enzyme reorganization, both at the level of conformational sampling and induced fit. Thus, a similar analysis cannot be performed with simple docking techniques. Effective alternative techniques are mostly limited to molecular dynamics simulations (Shan et al., 2011), although for the amount and type of studies that we will conduct it would require a significantly higher (almost prohibitive) computational cost.

More in detail, PELE combines the prediction of protein structure and Monte Carlo stochastic approach algorithms. The process begins with the calculation of the resting energy of the system, then an initial perturbation is generated. Afterward, the side-chain sampling is performed by an algorithm that places all side chains local to the ligand with a rotamer library side-chain optimization. The last step is the minimization with a truncated newton minimizer, which is a method for finding optimal points. These three steps give rise to a movement of the system which is tested and, if a new minimal is obtained, the movement will be valid and the simulation will continue. The process allows the efficient sampling of configurational space. PELE software can also simulate long processes very fast (**Figure 12**).

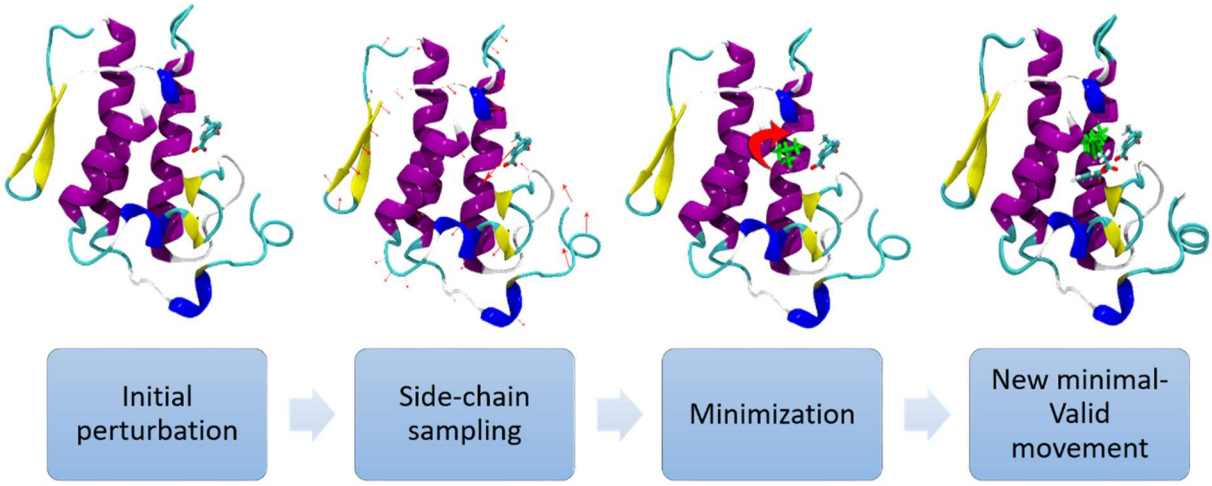


Figure 12. Scheme of PELE software workflow for modeling and mapping protein-ligand dynamics and induced fit in protein structures. Source: https://pele.bsc.es/pele.wt/

Remarks: Integration and deep analysis of biocatalytic data from an extensive set of enzymes will help to obtain an improved understanding, prediction, selection and design of novel commercially exploitable enzymes for industrially-driven applications. Here, molecular modeling and bioinformatic tools are considered to get a comprehensive understanding of the general molecular basis, if any, underlying what determines the enzyme properties, particularly substrate specificity.

Computational methods to engineer better enzymes based on predictive markers

The active site of an enzyme, which represents only a fraction of the total enzyme volume, has to include appropriate residues and/or a cofactor with a specific configuration, stabilizing substrates into reactive intermediates and transition states (Rufo et al., 2014). This combination of elements is the result of evolution from primitive peptide scaffolds (Toscano, Woycechowsky, & Hilvert, 2007; Rufo et al., 2014). With the discovery of the mechanisms by which proteins fold and the development of protein engineering and computational tools, it has been demonstrated that the active site and elements essential for stabilizing the transition states can be altered to modify catalytic properties. There are many examples in the literature where the application of rational protein design, site-directed or saturated mutagenesis and directed evolution have been used to predict and find key amino acid substitutions that designate a desired new or improved activity; for a recent example see Höhne et al. (2010) and Arnold (2019).

However, there is an important point to consider. All enzyme variants designed through rational design, molecular modeling or computational tools have been made in such way that the objective was to introduce (or alter) one-active site in a non-catalytic protein, probably because we mimic what commonly happens in nature: one enzyme – one active site (Frickel, Jemth, Widersten, & Mannervik, 2001; Deponte et al., 2007; Jiang et al., 2008). The reasons why a certain location in a protein is more favorable than others to establish an active site are totally unclear. Is it possible to find or design alternative sites to introduce an extra reactive group? No investigation has demonstrated that this is plausible; that is, no evidences have been gathered that demonstrate that introducing a second or more active sites into a polypeptide already containing a native one, is feasible. This demonstration was undertaken in this Thesis by using an approach consisting of locating extra binding pockets in an enzyme and turning them into catalytic active sites after appropriated mutations.

Remarks: As mentioned above, enzymes typically harbor one active site. Its properties can be redesigned, but the possibility to increase its number in an enzyme remains totally unexplored. In this Thesis, we attempt to evaluate whether the possibility to add additional active sites into a model enzyme, namely a serine ester-hydrolase from the structural superfamily of α/β-hydrolases, can help to promote the activity of the original enzyme. For that, we used the accumulated structural and biocatalytic meta-data and the predictive markers found relevant for defining substrate specificity.

Influencing flexibility constraints through immobilization

Like all molecules, proteins are capable to vibrate but as they are amino acid chains folded, they can also move as atom bonds can rotate. Different amino acids have different atoms and also a different number of them, so the global movement of a protein will depend on its primary structure, as well as its special distribution in the tertiary structure (**Figure 13**). This flexibility may determine most protein characteristics such as stability, activity, etc., as long as it represents the capability of responding to changes (Teilum, Olsen, & Kragelund, 2009).

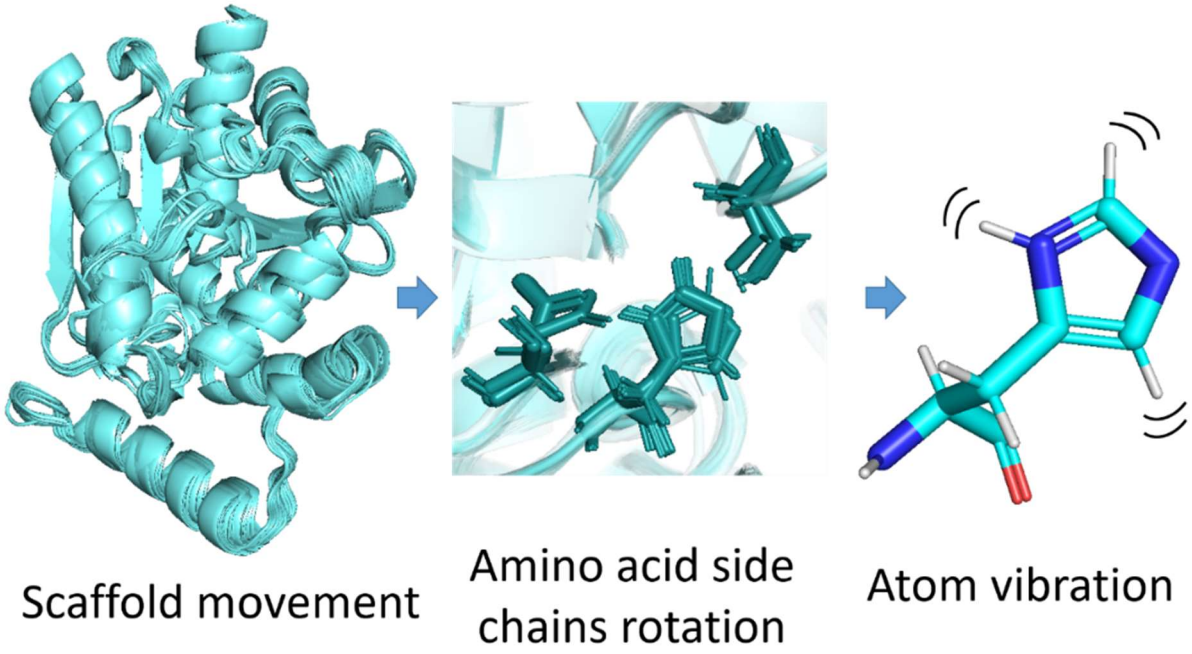


Figure 13. Schematic representation of movements conferring flexibility to proteins.

The main experimental techniques used to determine protein flexibility are spectroscopy and nuclear magnetic resonance (NMR) both of which require high specialization. However, although flexibility is an inherent property of a protein, it can also be modified, to an extent not known in many cases, through enzyme immobilization. Enzyme immobilization is a strategy of combining the enzyme with a carrier. It is widely used in the industrial processes for protecting the enzymes from the challenging condition used (Datta, Christena, & Rajaram, 2013). It consists on the confinement of an enzyme in a phase, which can be a matrix or a support, different from the one which contains substrates and products. Generally, there are 4 main groups of immobilization techniques which are summarized below and in **Figure 14**:

- Adsorption: n this immobilization protocol the enzyme is attached to the surface of the support by hydrophobic bonds and salt linkages. The adsorbed enzymes are protected from aggregation, proteolysis and interaction with hydrophobic interfaces.
- Covalent binding: This strategy is based on the creation of covalent bonds between the support and the side-chain of some amino acids of the protein such as lysine, arginine, aspartic acid and histidine. The support must be functionalized with some reactive group. In general, the covalent bonds give rise to a tight union that prevents the separation of the enzyme during utilization.
- Affinity immobilization: In this protocol, the binding of the enzyme can be achieved by two typical ways; in the first one a matrix is pre-coupled with an affinity ligand for the target enzyme, in the second, the enzyme is conjugated with some element that has some kind of affinity for the matrix.
- Entrapment: This method is based on the capture of enzymes by covalent or non-covalent bonds inside a polymeric matrix such as gels, fibers or pores. The entrapment prevents enzyme leakage and increases the mechanical stability.

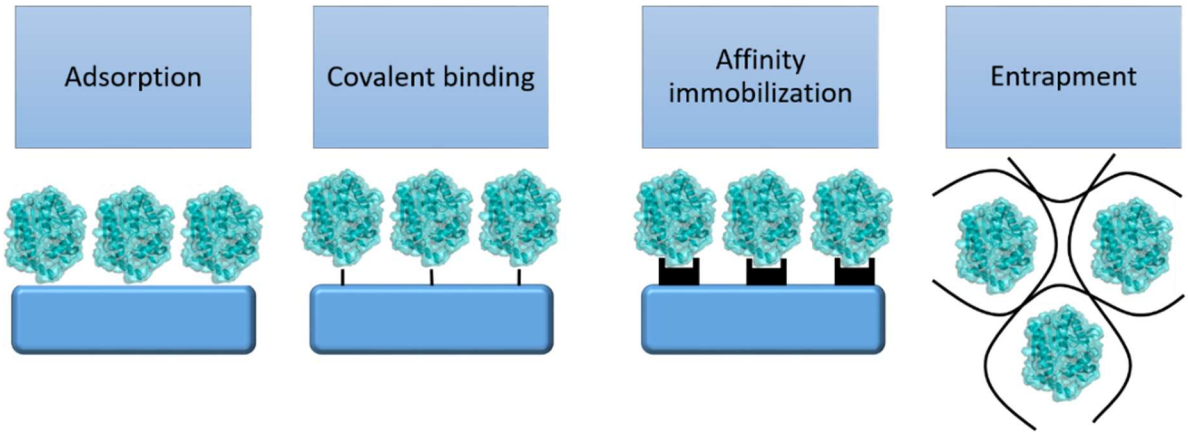


Figure 14. Scheme of main immobilization techniques.

As there are lots of immobilization protocols, the best one will depend on the enzyme selected for the process, and this is why a discussion on immobilization strategies and the outcomes will not be made in this Thesis. Having said that, it is to note that as soon as an enzyme is selected, the process to generate the final biocatalysts has to be worked out. The possible effects of the immobilization technique to be applied add significant complexity to the process, while at the same time opening a window to modify *a-la-carte* the enzyme properties. In this line, it is therefore critical to have a deep understanding of how enzyme immobilization influences substrate range. This is crucial since it would not help to find a promiscuous enzyme if its immobilization affects negatively this property. Opposite, through immobilization one may open the possibility to modulate the substrate range a step further.

Remarks: In this Thesis, different immobilization techniques are applied to a top promiscuous but not stereo-specific serine ester-hydrolase to evaluate the possibility of modulating its substrate spectra. The idea is not to stabilize the biocatalyst but rather deepen into the possibility of creating promiscuous and chiral-specific ones through supramolecular engineering.

In summary, it is clear that enzymes only end up in industrial processes if they meet at least a few criteria: high substrate range, high stereoselectivity and high turnover rates, if possible, under harsh reaction conditions. Major problems are related to the time-consuming and cost-intensive validation and process implementation of enzymes, which results in late market entry and substantial financial losses. Only very

few enzymes in current collections meet the criteria required for an industrial application. Narrow substrate spectra, low stereoselectivity and low-temperature tolerance are the most frequent problems. This implies that only a very small subset of proteins in an enzyme collection has properties that make them useful for a biocatalytic process. Imagine how many steps are needed to explore the environment, microbial genomes and metagenomes, for their enzymes: to organize a sampling expedition, the processing of samples, the discovery and also the selection and production of right ones, the validation, the quality assurance, the improvement through engineering, the pre-industrial and industrial studies and tests, the approval and the commercialization.

Based on the above, this Thesis fisrt focused on the creation of a unique and large collection of very important and probably amongst the most widely used class of enzymes for industrial biocatalytic applications, such as ester-hydrolases and transaminases. Sencondly, on the establishment of a pipeline to express such collection of enzymes in different host strains and plasmids and on purifying sufficient amounts for activity tests, substrate profiling and crystallization studies. Thirdly, on the integration of meta-biocatalytic data with solved crystal structures, 3D models, and computational information to understand and predict substrate specificity level and its relation with stereo-specificity, in both enzyme classes. Finally, this Thesis also comprises the establishment of a pipeline trained with the top candidate enzymes with engineering (genetic and supramolecular) tools to enable a swift identification and design of novel and robust enzymes and/or biocatalysts.

OBJECTIVES

The main scientific objective of this Thesis is to deliver an enzyme collection, which encompasses a high percentage of substrate promiscuous enzymes. The enzyme classes we strive to include are ester-hydrolases and class III ω-transaminases, which are amid the most important ones in biotechnology and biocatalysis. They can be implemented in the production of fine and platform chemicals, chiral building blocks and many other high-value compounds, and their application usually results in the speed-up of synthesis routes and thus a significant reduction of the environmental impact and time to market. Special attention is paid to enzymes accepting large and bulky substrates and, if possible, being stereo-specific. These general objectives are based on the following starting hypothesis: "enzymes with broad substrate range can be easily used in more than one real process to deliver different classes of "first in class" molecules in a cost- and time-effective way". Based on these considerations, the specific objectives of this Thesis are the following:

[1] To deliver, through metagenome and genome mining, a large collection of ester-hydrolases and class III ω-transaminases serving as model study targets to investigate substrate promiscuity.

[2] To use cutting-edge analytics and the enzyme collection to generate metadata, regarding substrate specificity and selectivity towards a broad range of substrates representing challenging chemical steps in some cases.

[3] To identify enzymes with broad substrate range and novel stereo-selectivity.

[4] To use structural, molecular modeling and computational tools to have a comprehensive understanding of the general molecular basis underlying what determines the properties of ester-hydrolases and class III ω-transaminases, particularly, broad substrate range, and later transfer these findings to other enzyme classes.

[5] To set up prediction tools to identify in databases, or create by rational design and supramolecular engineering, a large sub-collection of potentially marketable biocatalysts, with prominent substrate range while being stereo-specific, attractive enough for industrial process development.

METHODOLOGY, MATERIALS AND RESULTS

- Chapter 1: Determinants and prediction of esterase substrate promiscuity patterns.

- Chapter 2: Relationships between substrate promiscuity and chiral selectivity of esterases from phylogenetically and environmentally diverse microorganisms.

- Chapter 3: Bioprospecting reveals class III ω-transaminases converting bulky ketones and environmentally relevant polyamines.

- Chapter 4: Controlled manipulation of enzyme specificity through immobilization-induced flexibility constraints.

- Chapter 5: Rational engineering of multiple active sites in an ester-hydrolase.

Chapter 1: Determinants and prediction of esterase substrate promiscuity patterns.

This Chapter constitutes the central axis of this Thesis for four main reasons. First, because a large collection of 145 ester-hydrolases is created, through which we evaluate and understand the molecular basis of substrate promiscuity in this class of enzymes. Second, because the accumulated knowledge allows identifying ester-hydrolases with substrate specificity broader than that of best commercial preparations. Third, because the accumulated knowledge also allows establishing the relation between substrate promiscuity and enantio-selectivity, which will be discussed in Chapter 2. And finally, because by using as lead scaffold one of the enzymes of this study with the broadest substrate specificity we were able to generate biocatalysts with improved performances through supramolecular (Chapter 4) and protein (Chapter 5) engineering tools.

More in detail, in this Chapter we present an example of how an accumulated knowledge of enzyme´s biocatalytic properties can lead to the identification of molecular or structural marker of substrate promiscuity and how this marker can help accurately identifying substrate promiscuous enzymes in the databases, and to further produce and design biocatalysts for biotechnological purposes. An ester-hydrolase collection was created by applying genomic and metagenomic techniques and by establishing functional and *in silico* screenings. A total of 145 new and highly diverse ester-hydrolases were identified. For two of them, crystals were obtained and, for a sub-set of 96, 3D models were established. All of them were cloned, expressed and purified at mg scale, and after optimizing a high throughput assay, characterized against a wide library of 96 different chemically and structurally diverse substrates. The comprehensive analysis of the biocatalytic and structural data, complemented with computational analysis allowed the identification of the molecular determinants that can explain the higher level of substrate promiscuity that some enzymes show compared to others and how we can predict this property from a sequence. In particular, we described a structural parameter we called “effective volume”. This parameter shows the ratio between the total volume of the active site of the enzymes and the SASA (solvent accessible surface area) of the catalytic triad. In other words, the effective volume gives us a numerical value of how big and how exposed is the catalytic pocket of ester-hydrolases. We found this parameter, easily calculated from sequence information, being a marker of the promiscuity level, so that enzymes with an effective volume above a threshold do show a broader substrate profile than those being below the threshold. This marker was also found to be applicable to other enzyme classes such as phosphatases of the haloacid dehalogenase (HAD) superfamily of hydrolases. Therefore, through using this parameter one can screen databases for finding promiscuous ester-hydrolases, avoiding the steps of cloning and expression of enzymes encoded by all sequences.

The work also presents a set of ester-hydrolases converting more substrates, including industrially-relevant ones, than those of the best commercial preparations. Finally, this contribution also deepens into the microorganisms and environmental sites potentially containing enzymes with broad substrate range compared to others.

Note: In this article, I am the second author but with equal contribution as the first one. I have participated in all steps of the cloning, expression, purification and biocatalytic characterization of all 145 enzymes of the collection. I have also contributed to the design of the high throughput tools to evaluate substrate specificity and significantly contributed to the integration and analysis of the experimental, structural and computational data. The relatively high number of coauthors is due to the fact that such a great collection of enzymes could not be built without the participation of a high number of scientific groups, that have shared their genomic and metagenomic resources.

ACS chemical biology

Articles

Cite This: ACS Chem. Biol. 2018, 13, 225−234

Determinants and Prediction of Esterase Substrate Promiscuity Patterns

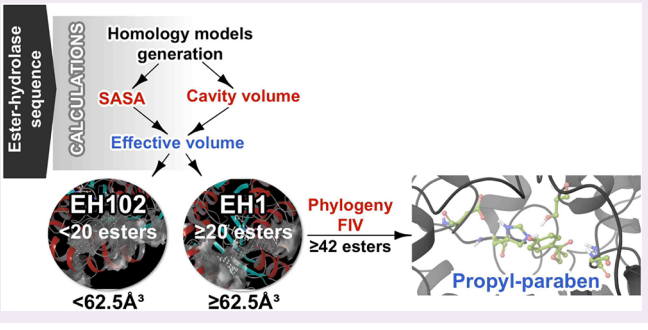
Monica Martínez-Martínez,[†,ζ] Cristina Coscolín,[†,ζ] Gerard Santiago,[‡,ζ] Jennifer Chow,[§] Peter J. Stogios,[∥] Rafael Bargiela,[†,γ] Christoph Gertler,[⊥,δ] Jose Navarro-Fernandez,[†] Alexander Bollinger,[#] Stephan Thies,[#] Celia Mendez-García,[▽,ϵ] Ana Popovic,[∥] Greg Brown,[∥] Tatyana N. Chernikova,[⊥] Antonio García-Moyano,[○] Gro E. K. Bjerga,[○] Pablo Perez-García,[§] Tran Hai,[⊥] Mercedes V. Del Pozo,[†] Runar Stokke,[◆] Ida H. Steen,[◆] Hong Cui,[∥] Xiaohui Xu,[∥] Boguslaw P. Nocek,[¶] María Alcaide,[†] Marco Distaso,[⊥] Victoria Mesa,[▽] Ana I. Pelaez,[▽] Jesus Sanchez,[▽] Patrick C. F. Buchholz,[%] Jurgen Pleiss,[%] Antonio Fernandez-Guerra,[$,●,■] Frank O. Glockner,[$,●] Olga V. Golyshina,[⊥] Michail M. Yakimov,[⌀,★] Alexei Savchenko,[∥] Karl-Erich Jaeger,[#,α] Alexander F. Yakunin,[∥,ψ] Wolfgang R. Streit,[§,ψ] Peter N. Golyshin,[⊥,ψ] Víctor Guallar,[*,‡,β,ψ] Manuel Ferrer,[*,†,ψ] and The INMARE Consortium

[†]Institute of Catalysis, Consejo Superior de Investigaciones Científicas, 28049 Madrid, Spain
[‡]Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain
[§]Biozentrum Klein Flottbek, Mikrobiologie & Biotechnologie, Universitat Hamburg, 22609 Hamburg, Germany
[∥]Department of Chemical Engineering and Applied Chemistry, University of Toronto, M5S 3E5 Toronto, Ontario, Canada
[⊥]School of Biological Sciences, Bangor University, LL57 2UW Bangor, United Kingdom
[#]Institut fur Molekulare Enzymtechnologie, Heinrich-Heine-Universitat Dusseldorf, 52425 Julich, Germany
[▽]Department of Functional Biology-IUBA, Universidad de Oviedo, 33006 Oviedo, Spain
[○]Uni Research AS, Center for Applied Biotechnology, 5006 Bergen, Norway
[◆]Department of Biology and KG Jebsen Centre for Deep Sea Research, University of Bergen, 5020 Bergen, Norway
[¶]Structural Biology Center, Biosciences Division, Argonne National Laboratory, Argonne, 60439 Illinois, United States
[%]Institute of Biochemistry and Technical Biochemistry, University of Stuttgart, 70569 Stuttgart, Germany
[$]Jacobs University Bremen gGmbH, Bremen, Germany
[●]Max Planck Institute for Marine Microbiology, 28359 Bremen, Germany
[■]University of Oxford, Oxford e-Research Centre, Oxford, United Kingdom
[⌀]Institute for Coastal Marine Environment, Consiglio Nazionale delle Ricerche, 98122 Messina, Italy
[★]Immanuel Kant Baltic Federal University, 236041 Kaliningrad, Russia
[α]Institute for Bio- and Geosciences IBG-1: Biotechnology, Forschunsgzentrum Julich GmbH, 52425 Julich, Germany
[β]Institucio Catalana de Recerca i Estudis Avancats (ICREA), 08010 Barcelona, Spain

Supporting Information

ABSTRACT: Esterases receive special attention because of their wide distribution in biological systems and environments and their importance for physiology and chemical synthesis. The prediction of esterases' substrate promiscuity level from sequence data and the molecular reasons why certain such enzymes are more promiscuous than others remain to be elucidated. This limits the surveillance of the sequence space for esterases potentially leading to new versatile biocatalysts and new insights into their role in cellular function. Here, we performed an extensive analysis of the substrate spectra of

continued...



Received: November 21, 2017
Accepted: November 28, 2017
Published: November 28, 2017

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

145 phylogenetically and environmentally diverse microbial esterases, when tested with 96 diverse esters. We determined the primary factors shaping their substrate range by analyzing substrate range patterns in combination with structural analysis and protein−ligand simulations. We found a structural parameter that helps rank (classify) the promiscuity level of esterases from sequence data at 94% accuracy. This parameter, the active site effective volume, exemplifies the topology of the catalytic environment by measuring the active site cavity volume corrected by the relative solvent accessible surface area (SASA) of the catalytic triad. Sequences encoding esterases with active site effective volumes (cavity volume/SASA) above a threshold show greater substrate spectra, which can be further extended in combination with phylogenetic data. This measure provides also a valuable tool for interrogating substrates capable of being converted. This measure, found to be transferred to phosphatases of the haloalkanoic acid dehalogenase superfamily and possibly other enzymatic systems, represents a powerful tool for low-cost bioprospecting for esterases with broad substrate ranges, in large scale sequence data sets.

Enzymes with outstanding properties in biological systems and the conditions favoring their positive selection are difficult to predict. One of these properties is substrate promiscuity, which typically refers to a broad substrate spectrum and acceptance of larger substrates. This phenomenon is important from environmental,[1] evolutionary,[2−5] structural,[6−8] and biotechnological[9,10] points of view. The relevance of substrate promiscuity is indisputable as the operating basis for biological processes and cell function. As an example, the evolutionary progress of enzymes from lower to higher substrate specificity allows the recruitment of alternate pathways for carbon cycling and innovations across metabolic subsystems and the tree of life by maximizing the growth rate and growth efficiency.[11] Promiscuous enzymes are energetically more favorable than specialized enzymes,[4] and therefore, the cell does not require many different enzymes to take up substrates, favoring genome minimization and streamlining.[12] In addition, the acquisition of new specificities without compromising primary or ancestral ones is a major driver of microbial adaptation to extreme habitats.[13] From a more practical standpoint, along with requirements of a technical nature such as selectivity, scalability and robustness, a narrow substrate spectrum is one of the most frequent problems for industrial enzyme applications.[14] A consensus exists that "the more substrates an enzyme converts the better," opening application ranges with consequent reduction of the production cost of multiple enzymes.[10,14,15]

Enzymes with wide substrate ranges occur naturally, as systematically investigated for halo-alkane dehalogenases,[16] phosphatases,[1] beta-lactamases,[2,17] and hydroxyl-nitrile lyases.[5] Some enzymes are more promiscuous than others simply due to their fold or degree of plasticity or the presence of structural elements or mutations occurring under selection in the proximity of the active-site cavity and access tunnels favoring promiscuity. However, the general explanation, if any, by which an enzyme binds and converts multiple substrates is unknown, although molecular insights have been reported for single enzymes.[18] A tool that can clearly distinguish promiscuous versus non-promiscuous enzymes and suggest substrates potentially being converted or not by them might therefore be valuable in applying low-cost sequencing in discovery platforms in any biological context.

In an ideal scenario, functional characterization of enzymes with genomics[19] and metagenomics[10,20] techniques using a large library of substrates would guide the analysis of sequence-to-promiscuity relationships and explore the mechanistic basis of promiscuity. In addition, such studies may help identify a new generation of highly promiscuous microbial biocatalysts. However, extensive bioprospecting and biochemical studies are rare,[10] despite the growing number of sequences available through low-cost sequencing efforts[21] and the growing number of enzymes that are typically characterized with limited substrate sets.[14] To address this knowledge gap, we functionally assessed the substrate specificity of a set of 145 phylogenetically, environmentally, and structurally diverse microbial esterases (herein referred to as "EH," which means Ester Hydrolase) against a customized library of 96 different substrates to find predictive markers of substrate promiscuity rather than discrete determinants of substrate specificity that may differ from protein to protein. EHs were selected for an analysis of substrate promiscuity because they typically have specific definitions of molecular function, can be easily screened in genomes and metagenomes compared with many other classes of proteins, are among the most important groups of biocatalysts for chemical synthesis, and are widely distributed in nature, with at least one EH per genome.[14]

Our work adds important insights and empirical, structural, and computational data to facilitate the elucidation of the molecular basis of substrate promiscuity in EHs, which was further extended to phosphatases from the haloalkanoic acid dehalogenase (HAD) superfamily. This was achieved by deciphering what we consider a predictive structural marker of substrate promiscuity and by establishing the reasons why certain such enzymes are more promiscuous than others and can convert substrates that others cannot. This study does not pretend to generate a quantitative measure to predict the number of compounds that an enzyme will hydrolyze but a tool and a parameter that will help in ranking (classifying) promiscuity level. Following on from that, we propose in this work the first molecular classification method of this kind derived from first principle molecular simulations and with clear physical/structural interpretation. This work also provides an example of the utility of this parameter to screen the sequence space for highly promiscuous EHs that may compete with best commercial EH preparations. We also provide first preliminary evidence of a number of underexplored microbial phylogenetic lineages containing EHs with a prominent substrate range.

RESULTS AND DISCUSSION

The Substrate Range of 145 Diverse EHs. A total of 145 EHs were investigated. Extensive details of the sources and screen methods are provided in the Supporting Information Methods and Table S1. In an environmental context, the source of enzymes was highly diverse because they were isolated from bacteria from 28 geographically distinct sites (125 EHs in total) and from six marine bacterial genomes (20 EHs; Supporting Information Figure S1). A phylogenetic analysis also indicated that sequences belong to bacteria distributed across the entire phylogenetic tree (Supporting Information Results and Figure S2).

The 145 putative proteins exhibited maximum amino acid sequence identities (Supporting Information Table S1) ranging from 29.1 to 99.9% to uncharacterized homologous proteins in public databases, with an average value (reported as %, with the

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

interquartile range (IQR) in parentheses) of 74.3% (40.3%). The pairwise amino acid sequence identity for all EHs ranged from 0.2 to 99.7% (Supporting Information Table S2), with an average value of 13.7% (7.6%). BLAST searches were performed for all query sequences by running NCBI BLASTP against the current version of the Lipase Engineering Database[22] using an E-value threshold of 10^{-10} and were successful for all but nine candidates. A total of 120 EH sequences were unambiguously assigned to some of the 14 existing families (F) of the Arpigny and Jaeger classification, which are defined based on amino acid sequence similarity and the presence of specific sequence motifs.[14,23] These EHs included sequences with a typical α/β hydrolase fold and conserved G-X-S-X-G (FI, 20; FIV, 36; FV, 33; FVI, 5; and FVII, 6) or G-X-S-(L) (FII, 9) motifs and sequences with a serine beta-lactamase-like modular (non-α/β hydrolase fold) architecture and a conserved S-X-X-K motif (FVIII, 11). An additional set of nine sequences were assigned to the *meta*-cleavage product (MCP) hydrolase family[24] and six to the so-called carbohydrate esterase family,[25] both with typical α/β hydrolase folds. Finally, one was a cyclase-like protein from the amido-hydrolase superfamily.[26] Sequences-to-family assignments are summarized in the Supporting Information Table S1. Taken together, the primary sequence analysis suggests that the diversity of polypeptides is not dominated by a particular type of protein or highly similar protein clusters but consists of diverse nonredundant sequences assigned to multiple folds and subfamilies, which are distantly related to known homologues in many cases.

The substrate profiles of all EHs were examined using a set of 96 chemically and structurally distinct esters (Supporting Information Table S3). We are aware that the number of compounds hydrolyzed may be an ambiguous indicator of promiscuity, because the size and composition of the library may influence the results. For this reason, the composition of the library was not random but based on including esters with variation in size of acyl and alcohol groups and with growing residues (aromatic, aliphatic, branched, and unbranched) at both sides, leading to more challenging substrates because a larger group adjacent to the ester bond increases the difficulty of conversion. Halogenated, chiral, and sugar esters, lactones, and an alkyl diester were also included. Esters with nitro substituents were not included. We used the partitioning coefficient (log P value) to indicate the chemical variability of the esters because this parameter reflects electronic and steric effects and hydrophobic and hydrophilic characteristics. Log P was determined with the software ACD/ChemSketch 2015.2.5. Log P values (Supporting Information, Table S3) ranged from −1.07 (for methyl glycolate) to 23.71 (for triolein), with an average value (IQR in parentheses) of 3.13 (2.86), which indicates that the ester library used in this study had broad chemical and structural variability. Nevertheless, adding new substrates could surely help (and even change) the ranking of the EHs herein analyzed. The dynamic range of the assay may also influence the results. For this reason, to detect enzyme−substrate pairs for a given EH, the ester library was screened with each of the 145 EHs in a kinetic pH indicator assay in 384-well plates,[24,27,28] which unambiguously allow quantifying specific activities at pH 8.0 and 30 °C, using a substrate concentration above 0.5 mM (see Supporting Information, Results). Two commercial lipases, CalA and CalB from *Pseudozyma aphidis* (formerly *Candida antarctica*), were included in the assays for comparison. Using this data set, we linked the biocatalytic data to the sequence information for the respective enzyme. In this study, sequence information meant any sequence that encoded an EH of interest. Biocatalytic data meant experimental data on substrate conversion (i.e., units g^{-1} or U g^{-1}) followed for 24 h.

We determined the probability of finding an EH with a broad substrate profile by plotting the number of esters that were hydrolyzed by all preparations. Figure 1 shows that the number

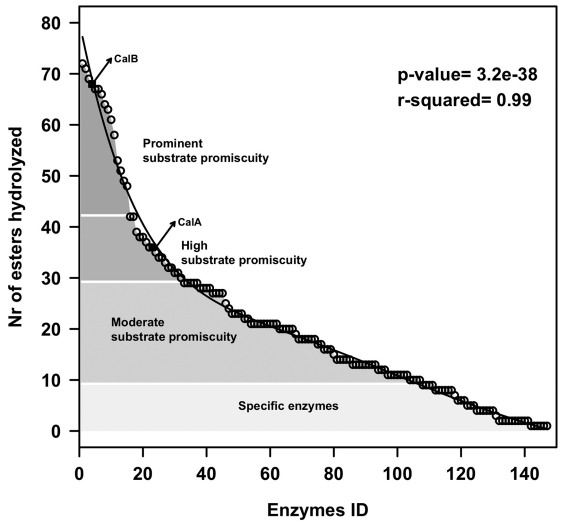


Figure 1. Number of ester substrates hydrolyzed by each of the 145 EHs investigated in this study. The commercial preparations CalA and CalB (marked with filled square) are also included. This figure is created from data in the Supporting Information Table S1. The activity protocol established and used to identify the esters hydrolyzed by each EH was based on a 550 nm follow-up pH indicator assay described in the Supporting Information Methods. The list of the 96 structurally different esters tested is shown in Figure 2. Full details of the activity protocol are provided in the Supporting Information Methods. The trend line shows a not-single exponential fit of the experimental data. The fit was obtained using R script and the "lm" function, to extract a polynomial regression of degree 6 with the following line "model ← lm(MM[,1] ~ poly(MM[,2],6,raw = TRUE))", where MM[,1] corresponds to the number of esters hydrolyzed, and MM[,2] the position in the *x* axis (from 1 to 147).

of esters hydrolyzed by all 147 EHs (including CalA/B) fits to an exponential distribution (r^2 = 0.99; p value $3.2e^{-38}$; Pearson's correlation coefficient) with a median of 18 substrates per enzyme, nine hits at the 25th percentile, and 29 hits at the 75th percentile. On the basis of this distribution and a previously established criterion,[1] we considered an enzyme specific if it used nine esters or fewer (27% of the total), as showing moderate substrate promiscuity if it used between 10 and 29 esters (51% of the total), and as showing high-to-prominent promiscuity if it used 30 or more esters (22% of the total). This criterion indicated a percentage of EHs with a prominent substrate range similar to that found for HAD phosphatases (24%).[1]

Phylogeny Is a Predictive Marker of Substrate Promiscuity. Hierarchical clustering was performed to evaluate the differences in substrate range patterns (Figure 2). For the sake of simplicity, clustering was performed for those EHs that hydrolyzed 10 or more esters (i.e., 107 total EHs). We first observed a large percentage of enzymes with presumptive broad active site environments that accommodated large aromatic and sterically hindered esters such as benzyl (*R*)-(+)-2-hydroxy-3-phenylpropionate (49% of the total), benzoic acid-4-formyl-phenylmethyl ester (27%), 2,4-dichlorophenyl 2,4-dichlorobenzoate (~8%), 2,4-dichlorophenyl 2,4-dichlorobenzoate (~5%), and diethyl-2,6-dimethyl 4-phenyl-1,4-dihydro

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

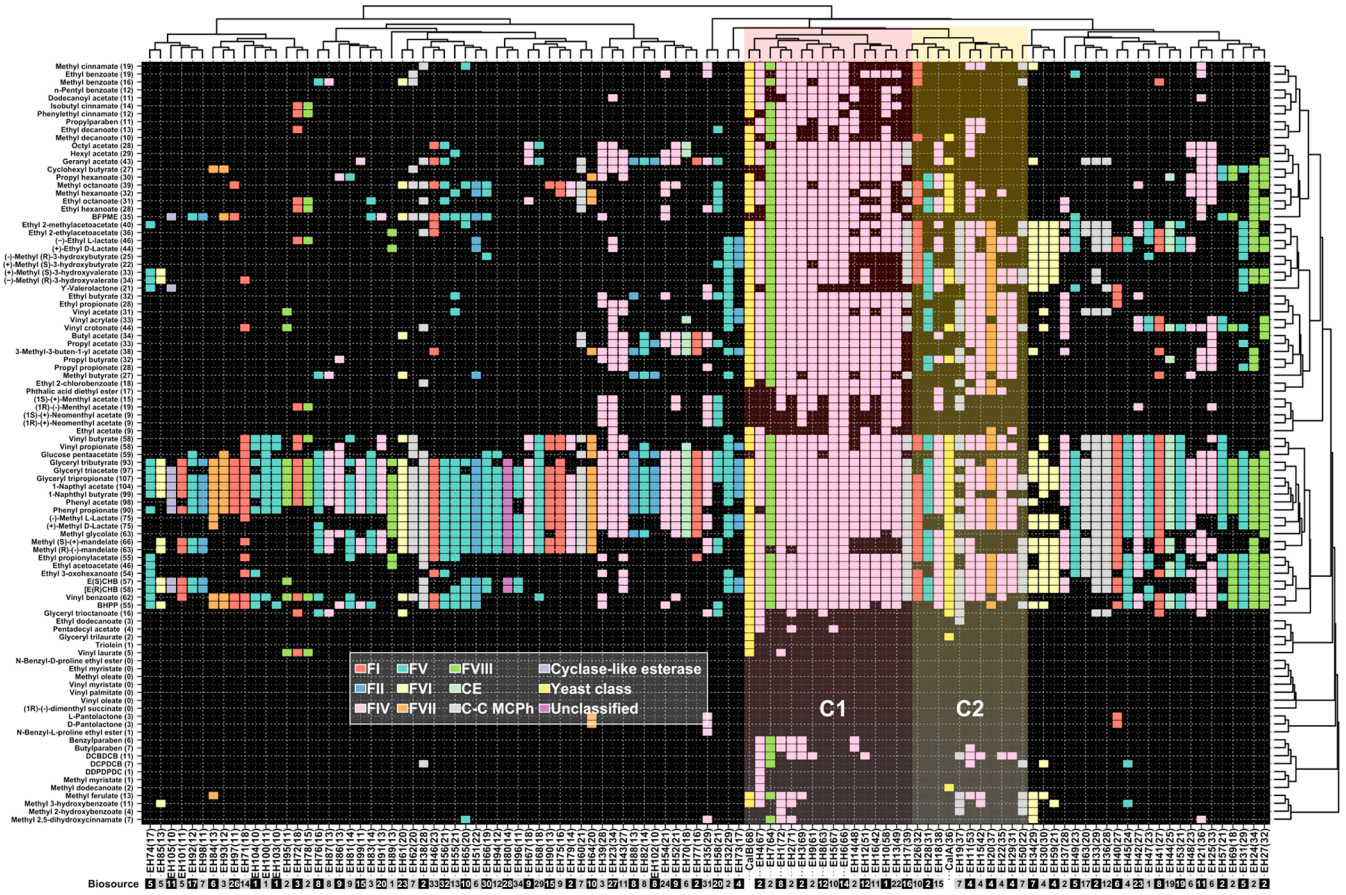


Figure 2. Hierarchical clustering of the substrate ranges of the EHs. Only EHs that hydrolyzed 10 or more esters were considered (107 in total, including CalA/B). This figure is created from data in the Supporting Information Table S3. The specific activities of the EHs for each of the 96 esters were determined as described in Figure 1. The list of the 96 esters tested and the frequency of each ester considered as a hit (in brackets) are shown on the left side. The ID code representing each EH is given at the bottom. Each hydrolase is named based on the code "EH," which means Ester Hydrolase, followed by an arbitrary number from 1 to 145 for the most to least promiscuous enzyme. The number in brackets indicates the number of esters hydrolyzed by each enzyme. The biosource of each EH is indicated at the bottom with a number in white or black squares that follows the nomenclature in the Supporting Information Figure S1. The figure was created with the R language console using a binomial table with information about the activity/inactivity (1/0) of the analyzed enzymes against the 96 substrates as a starting point. For the central graphic, which shows the data in Supporting Information Table S3, we used the drawing tools provided by the basic core packages of R. The hierarchical clusters of the enzymes (shown at the top) and substrates (shown on the right side) were generated by calculating a distance matrix using a "binomial" method and the hclust function to generate the tree. Using the functions as .phylo and plot.phylo from the ape package, the clusters were added to the top and right of the figure. A combination of the Set1 palette from the R package RColorBrewer and colors from the basic palette from R were used as the color palette for sequences assigned to each family (F; see inset), including FI to FVII, carbohydrate esterase (CE), and carbon−carbon *meta*-cleavage product hydrolase (C−C MCP) families, all with a typical α/β hydrolase fold, FVIII serine beta-lactamase with non α/β hydrolase fold, and cyclase-like protein from the amido-hydrolase superfamily. Sequences that were not unambiguously ascribed to existing families were referred to as "Unclassified," and those of yeast origin were assigned to "yeast class." The two "clusters" C1 and C2 that contained the most substrate-promiscuous EHs are color-coded under a shadowed background.

pyridine-3,5-dicarboxylate (~1%). Therefore, even though the EHs in this study were identified by a selection process based on the utilization of short esters (see Supporting Information Methods), the isolation of EHs with ample substrate spectra and the ability to hydrolyze very large substrates was not compromised.

We detected drastic shifts in substrate specificity (Figure 2), with glyceryl tripropionate as the only substrate hydrolyzed by all EHs. This is consistent with the high sequence variability within EHs, with an average pairwise identity of 13.74%. We then sought to determine the primary factors shaping the substrate range and thus defined different functional clusters. First, we observed that global sequence identity was of limited relevance for inferring the substrate range because no correlation was found (r^2 = 0.25) between the differences in identity and the number of esters that were hydrolyzed (Supporting Information Tables S1 and S2). Second, comparisons of the substrate range and the hydrolysis rate (U g^{-1} for the best substrates) were performed (Supporting Information Table S1). No correlation existed (r^2 = 0.073), suggesting that our assay conditions allow evaluating the promiscuity level whatever the hydrolytic rate of the EH is. In addition to the low correlation values, no threshold above or below which one could qualitatively classify the substrate range was observed in both cases, so that sequence identity and hydrolytic rate are neither predictive nor classification parameters of promiscuity. Additionally, no link between substrate range and habitat was found because EHs from the same biosource fell into separate clusters (Figure 2). Phylogeny-substrate spectrum relationships were further examined. Figure 2 indicates that the broad substrate-spectrum EHs did not cluster in a single phylogenetic branch, yet substrate promiscuity was mostly found for members of one of 10 subfamilies covered. Indeed, 67% of the EHs that could hydrolyze 30 or more esters

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

(mostly located in clusters C1 and C2 in Figure 2) were assigned to FIV,[14,23] and this percentage increased to 84% when considering only those EHs that could hydrolyze 42 to 72 esters (Figure 2; cluster C1). In addition to FIV members, a FVIII serine beta-lactamase showed prominent substrate spectra (see cluster 1). Members of both families (FIV, 8; FVIII, 1; see cluster C1) hydrolyzed as many esters (from 61 to 72) as the yeast family member CalB (68 esters), the most promiscuous commercially available lipase preparation used for the production of fine chemicals.[29]

Phylogeny was thus indicated as a predictive marker of the substrate range of EHs, as although a broad substrate scope was assigned to several sequence clusters, this feature was prevalent in members of FIV. A query sequence that matched FIV could be easily identified by means of the consensus motif GDSAGG around the catalytic serine; this family is also called the hormone-sensitive lipase (HSL) family because a number of FIV EHs display a striking similarity to the mammalian HSL.[14,23] Noticeably, the location of some FIV members in functional clusters with narrow substrate spectra (Figure 2) suggests that factors other than phylogeny contribute to the substrate spectra of EHs.

The Active Site Effective Volume Is a Prominent Marker of EH Promiscuity. Structural-to-substrate spectrum relationships were further examined by protein−ligand simulations to find additional markers of promiscuity. Crystals from recombinant EH1,[28] the protein with the broadest substrate range under our assay conditions, were obtained as described in the Supporting Information Methods. The enzyme with the widest substrate range was considered the best candidate for understanding the nature of promiscuity. This enzyme seems to have a wide active site environment as, under our assay conditions, it accepted 72 esters ranging from short (*e.g.*, vinyl acetate) to large (*e.g.*, 2,4-dichlorobenzyl-2,4-dichlorobenzoate; Figure 2). We also obtained crystals of recombinant EH102, which was isolated from the same habitat[28] but had a restricted substrate range, hydrolyzing only 10 of the 96 esters tested (Figure 2). Crystallographic data and refinement statistics for the two structures are given in Supporting Information Table S4.

To rationalize the substrate range shown by EH1 and EH102, we performed substrate migration studies using the software Protein Energy Landscape Exploration (PELE), which is an excellent tool to map ligand migration and binding, as shown in studies with diverse applications.[30−32] To map the tendency of a substrate to remain close to the catalytic triad, the substrate was placed in a catalytic position, within a proton abstraction distance from the catalytic serine, and allowed to freely explore the exit from the active site. The PELE results for both proteins and glyceryl triacetate are shown in Figure 3a. Clearly, EH1 has a significantly better binding profile, with an overall lower binding energy and a better funnel shape, whereas EH102 had a qualitatively unproductive binding-energy profile. This difference in the binding mechanism can be explained by the catalytic triad environment. EH1 has a somewhat wide but buried active site, whereas EH102 has a surface-exposed catalytic triad (Figure 4a). These structural differences translate into significant changes in the active site volume, as defined using Fpocket; the active site cavity of EH1 is 3-fold larger than that of EH102. Moreover, important changes are observed when inspecting the solvent exposure of the cavity. Figure 3b shows the relative solvent accessible surface area (SASA) for the substrate along the exploration of PELE, computed as a (dimensionless) percentage (0−1) of the ligand SASA in solution. Even at catalytic positions (distance Ser(O)−substrate(C) ∼ 3−4 Å), in EH102 we observe that ∼40% of the surface of the substrate is accessible to the solvent, which greatly destabilizes the substrate and facilitates escape to the bulk solvent. By contrast, EH1 has a larger but almost fully occluded site, with relative SASA values of approximately 0−10%, which can better stabilize the substrate.

After defining key points underlying the promiscuity of EH1, i.e., a larger active site volume and a lower SASA (Figure 4a), we extended the analysis to other EHs. First, we collected all 11 available crystal structures (Supporting Information Table S1) and computed the active site volume and relative SASA of the catalytic triad (Figure 5, square symbols). We next extended the analysis to the rest of the EHs using homology modeling (using the 11 crystals available) and produced a structural model for 84 additional enzymes. The missing ones were those with sequence identities of less than 25% (to an existing crystal) or those for which the catalytic triad could not be unambiguously identified (*i.e.*, not suitable alignments). Figure 5 (circle symbols) shows the active site effective volume data for all structural models. The analysis indicated a ratio threshold of 62.5 Å^3 for qualitatively classifying substrate promiscuity. Note that the relative SASA of the catalytic triad (derived from the GetArea server, see Supporting Information Methods) adopts values of 0−100; the actual value of the effective volume threshold will depend on the chosen range. We observed that values equal to or higher

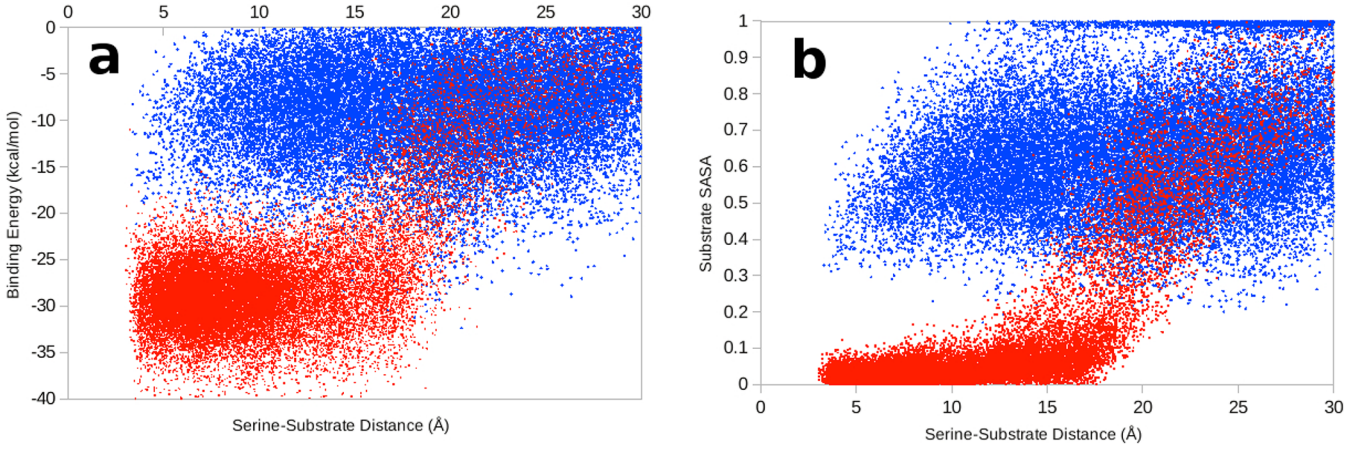


Figure 3. Protein Energy Landscape Exploration (PELE) analysis. Panel a shows the protein−substrate interaction plots for EH1 (red) and EH102 (blue). Panel b shows the relative SASA for glyceryl triacetate in EH1 (red) and EH102 (blue) computed as a dimensionless ratio (0−1) using PELE.

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

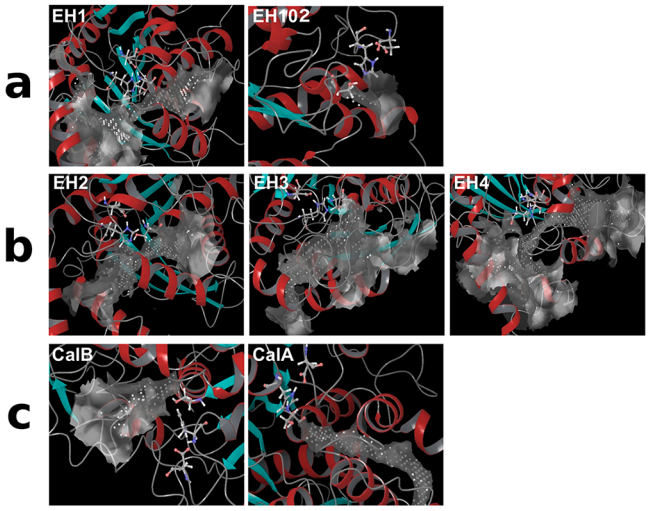


Figure 4. Catalytic triad exposure of selected EHs with the broadest and lowest substrate ranges. (a) The catalytic triad (ball-and-sticks) and the main adjacent cavity (gray clouds) as detected by SiteMap are underlined to demonstrate the differences between a promiscuous (EH1) and nonpromiscuous (EH102) EHs. EH1 can hydrolyze 72 esters and has a defined hidden binding cavity (effective volume: 166.7 Å^3). EH102, by contrast, can hydrolyze only 10 esters and has a surface-exposed triad (high SASA) and an almost negligible binding cavity (38.5 Å^3). The three top EHs with the broadest substrate ranges (b), positioned in the ranking after EH1, and the commercial CalB and CalA lipases (c), are also represented. On each panel, we highlight the catalytic triad and the main adjacent cavity as detected by SiteMap, demonstrating the differences in active site topology. EH2, EH3, and EH4, all assigned to FIV (as EH1), hydrolyzed 71, 69, and 67 esters and have defined but distinct hidden binding cavities (500, 200, and 200 Å^3, in the same order), as EH1. CalB, which was capable of hydrolyzing 68 esters, has a binding cavity (200 Å^3) that is also hidden but highly different from those of the other EHs. CalA, by contrast, hydrolyzed only 36 esters and has a low surface-exposed triad (SASA), with restrictive access to the catalytic triad (1000 Å^3).

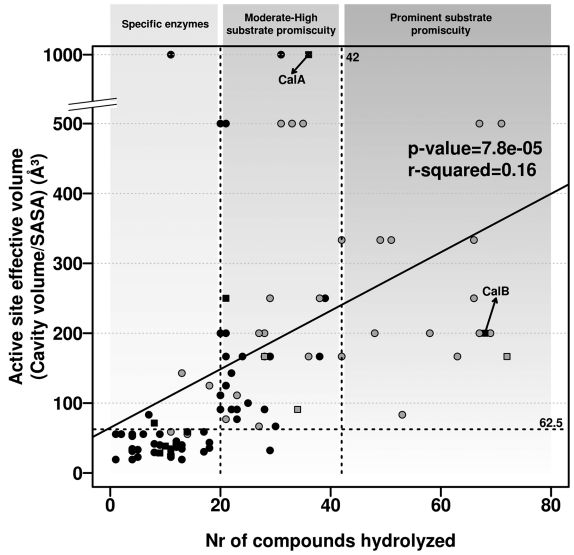


Figure 5. Defining of the substrate range of the EH by topology of the catalytic environment. The figure shows the relationships between the active site effective volume (in Å^3) and enzyme promiscuity (number of substrates hydrolyzed). Note that the presented data were obtained using the active site cavity volume computed in Å^3 and SASA as a dimensionless ratio from 0 to 100 using the GetArea server (http://curie.utmb.edu/getarea.html). The panel contains information for EHs for which crystal structures (square) and homology models (circles) could be unambiguously established (sequence identity ≥25%) and the catalytic triad identified. Gray circles and squares indicate the EHs assigned to FIV. The analysis indicated a threshold ratio (indicated by a horizontal dashed gray line) at which it is possible to qualitatively classify substrate promiscuity based on hydrolysis of at least 20 substrates. Phylogenetic analysis further extended the substrate spectra to ≥42 esters, as only enzymes assigned to FIV and conforming to the 62.5 Å^3 threshold, together with CalB, were capable of converting such a high number of esters. The positioning for the commercial CalA and CalB lipases is indicated.

than 62.5 Å^3 corresponded to EHs with activity for 20 or more of the 96 substrates tested and the opposite. There were only six outliers out of 95 EHs that did not follow this rule. Thus, the performance is of excellent (with 94%) accuracy if used as a classifier. The effective volume, however, does not have quantitative predictions for the exact number of esters hydrolyzed (r^2 = 0.16 for data in Figure 5), most likely because above the 62.5 Å^3 threshold, the capability to hydrolyze more or less substrates may specifically depend on the topology of the catalytic environment (Figure 4a–c), which may differ within families. Particularly, none of the different family members that conformed to the ≥62.5 Å^3 threshold, except those from FIV (*i.e.*, at least 50% of its members as shown in Figure 5, gray circle symbols) and CalB, could hydrolyze 42 or more esters. Therefore, the classification potential of the effective volume measure increased when combined with phylogenetic data. Noticeably, we observed that the predictive capacity of cavity volume/SASA is not influenced by the presence of flexible elements in the structure (Supporting Information Results).

The Active Site Effective Volume Is Also Indicative of Molecules Being Accepted As Substrates. We further used the active site cavity volume/SASA to also dissect its role in substrate specificity. We restricted the analysis to the 96 EHs for which this value could be unambiguously calculated (see above). The analysis indicated that the conversion of 34 esters was only observed for EHs conforming to the ≥62.5 Å^3 threshold (Supporting Information Figure S3). All but two (vinyl crotonate and ethyl acetate) could be considered large alkyl or hindered aromatic esters and included important molecules in synthetic organic chemistry such as paraben esters. This suggests that active sites with larger volume and a lower SASA (i.e., cavity less exposed to the surface) will most likely support hydrolysis of these esters. Therefore, the effective volume measure could be used to some extent as an indicator of substrates that may or may not be hydrolyzed by EHs. However, not all EHs fitting the ≥62.5 Å^3 threshold could convert all 34 of these esters, implying that this measure does not allow deepening into substrate specificity, which may depend on the topology of the catalytic environments as mentioned previously (Figure 4a–c). However, we found that the probability that benzyl-, butyl-, and propyl-paraben esters, major intermediates in chemical synthesis, are converted by members of the FIV with an effective volume ≥ 62.5 Å^3 is significantly higher (~35%) than that of EHs from FIV with a volume < 62.5 Å^3 and EHs from other families, whatever the value of the effective volume (approaching zero percent in our study); for those EHs for which effective volume could not be calculated, this probability is as low as 1.9% (Supporting Information Figure S4). This again exemplifies that the effective volume measure, when combined with phylogenetic information, is not only indicative of a promiscuity level but also can be used to predict the capacity to hydrolyze esters such as paraben esters. Screen programs to find EHs capable of converting

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225–234

paraben esters should most likely be directed to find those assigned to FIV and with cavity volume/SASA ≥ 62.5 Å^3.

The Effective Volume Is Also a Marker of Substrate Promiscuity in Proteins Other than EHs. In order to evaluate the possibility that the active site effective volume may be a marker of substrate promiscuity in other enzymes, substrate spectra-effective volume relationships should be investigated in other protein families. In this line, Huang *et al.*[1] recently performed a systematic analysis of the substrate spectra of 200 phosphatases of the HAD superfamily, when tested against a set of 167 substrates. We collected the available crystal structures of each of the HAD phosphatases (Supporting Information Table S5) and computed the active site effective volume. We restricted the analysis to C2 cap members as they were reported to have a broader substrate spectrum,[1] and crystal structures with low to high effective volume are available. Interestingly, we observed that the effective volume (using the two conserved aspartic catalytic residues as the corrective SASA factor) was highly correlated (r^2 = 0.92) with the substrate range (Figure 6). Thus, the effective volume can be used as a molecular classification parameter of substrate promiscuity of phosphatases of the HAD superfamily when crystal structures are available. When this analysis was extended to the rest of the enzymes using homology modeling, we observed a similar trend to that of EHs (Supporting Information Figure S5). That is, no correlation existed (r^2 = 0.043), but still the effective volume can be used as a classifier of the substrate range as for EHs. Indeed, although a threshold could not be unambiguously established, sequences with the top 10 effective volumes belong to moderate-high to high promiscuity enzymes.

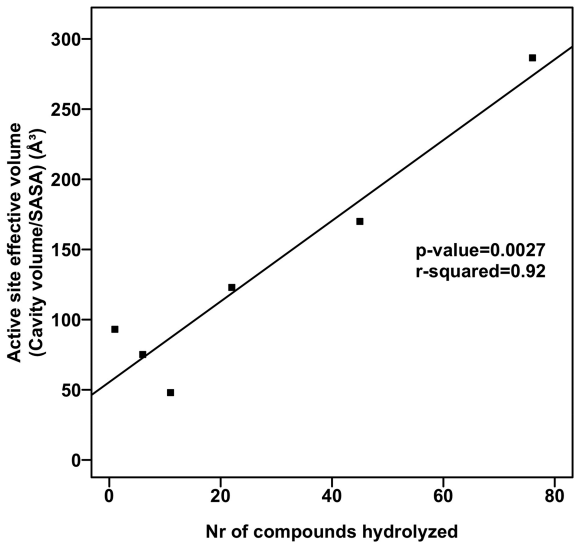


Figure 6. Relationships between the active site effective volume (in Å^3) and enzyme promiscuity (number of substrates hydrolyzed) of C2 members of HAD phosphatases. The number of substrates converted by each HAD phosphatase was obtained from Huang *et al.*[1] and is summarized in Supporting Information Table S5. The panels contain information for HAD phosphatases for which crystal structures were available and the catalytic residues identified. The active site effective volume (in Å^3) was calculated as described in Figure 5.

In conclusion, we found that the topology around the catalytic position, by means of an active site effective volume (cavity volume/SASA) threshold, is a dominant criterion of substrate promiscuity in EHs, which can be further extended by adding phylogenetic analysis. The rationale behind this parameter is as follows. Large volumes increase promiscuity until a certain value at which the cavity becomes too exposed and is not capable of properly accommodating and, importantly, retaining the substrate in specific catalytic binding interactions. This point is well captured by the SASA percentage of the catalytic triad, a dimensionless ratio that corrects for large volume measures in exposed sites. Importantly, the parameters of active site volume and relative SASA can be easily transferred to other systems. Indeed, the fact that the EHs investigated herein have different folds and that this parameter was also a marker of substrate spectra for phosphatases of the HAD superfamily opens the possibility of applying the effective volume measure to other enzymes requiring substrate anchoring. In all cases, the effective volume threshold-to-substrate relationships must be established. We would like to make note that the active site volume is not a static property, as the active site will breathe, depending on how flexible the protein is. In addition to that, the 62.5 Å^3 threshold for qualitatively classifying substrate promiscuity is based on the analysis of 147 EHs when tested against 96 esters. Although increasing the number of EHs and esters may influence this threshold and increase accuracy, it will not affect the fact that the measurement of the effective volume (cavity volume/SASA) can be used as the first molecular classification method of substrate promiscuity in EHs.

Our measurement is not a quantitative one, but rather a qualitative ranking (classification) procedure that will allow, for example, selecting sequences in databases for expression, particularly those encoding promiscuous enzymes capable of converting multiple substrates. This will substantially reduce reagent and labor costs compared to methods requiring the extensive cloning of all genes, and the expression and characterization of all enzymes in databases to later find those being promiscuous.[33] This possibility was herein examined by successfully mapping the open reading frames from the TARA Oceans project assemblies[34] and by identifying a high number of sequences encoding EHs with presumptive prominent substrate promiscuity (Supporting Information Results, Figures S6 and S7). Application of the effective volume measure to examine the sequences daily generated or deposited in databases requires having some crystals or X-ray structures for the model production. This limitation prevents predicting promiscuity from sequences lacking any structural information. Indeed, 36% of the EHs in this study (52 of the 147, including CalA/B) could not be included in the correlation because no calculation was possible. Accumulation of structural information and design and application of better modeling algorithms in the future will help solving this limitation.[35] Future studies might also explore molecular dynamics (MD) simulations to measure also the flexibility of the active site and not just the size of the cavity. By using this strategy, it was recently reported that the broad promiscuity of the members of the alkaline phosphatase superfamily arises from cooperative electrostatic interactions in the active site, allowing each enzyme to adapt to the electrostatic needs of different substrates.[36] In the particular case of EH phylogeny, a marker which does not require a three-dimensional structure was also suggested as a predictive classification marker of the substrate range. Indeed, this study suggests that in case of an unknown EH for which a crystal structure is not available or a homology model could not be established, then its assignation to family IV[14,23] increases the likelihood that this EH is promiscuous.

The present study not only provides clear evidence that substrate promiscuity in EHs has evolved from different core structural domains fitting an effective volume around the active

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

site, albeit with a bias toward that occurring in FIV members, but also from different phylogenetic lineages, many of which remain underexplored to date (Supporting Information Results and Figure S2). These are new findings as it was previously thought that the substrate range in a superfamily increased from a single ancestral core domain,[1] and because the identities of some microbial groups containing promiscuous enzymes, herein EHs, were previously unknown. Finally, this study also enabled the selection of a set of EH candidates that can compete with best commercial EHs such as CalB, as they show a broader substrate profile and specific activities up to 3-fold higher (Supporting Information Table S6). Their sequences can be used to search databases for similar promiscuous EHs. Further investigations should also determine the occurrence of other types of promiscuous EH phenotypes with broader substrate ranges than those identified in this study. For example, at least the stability of substrate-promiscuous EHs at different temperatures and with various solvents, along with the occurrence and evolution of secondary reactions, should be investigated in terms of condition and catalytic promiscuity.

METHODS

Protein Samples. Two main sources of EHs were used in the present study, all of them isolated *via* naive and sequence-based screens in genomes and metagenomes. A first set of samples was EHs previously reported, as in the bibliography (69 in total), and that were herein substrate-profiled for first time. A second set was EHs (77) that are herein reported for first time. The extensive details of the source, cloning, expression, and purification of each of the active and soluble EHs are provided in the Supporting Information Methods and Table S1.

Ester Bond Hydrolysis Activity Assessment: Substrate Profiling Tests with 96 Esters. Hydrolytic activity was assayed at 550 nm using 96 structurally diverse esters in 384-well plates as previously described.[24,27,28] Before the assay, a concentrated stock solution of the esters was prepared at a concentration of 100 mg mL^{-1} in acetonitrile and dimethyl sulfoxide (DMSO). The assays were conducted according to the following steps. First, a 384-well plate (Molecular Devices, LLC, CA, USA) was filled with 20 μL of 5 mM *N*-(2-hydroxyethyl)piperazine-*N*′-(3-propanesulfonic acid (EPPS) buffer, at pH 8.0, using a QFill3 microplate filler (Molecular Devices, LLC, CA, USA). Second, 2 μL of each ester stock solution was added to each well using a PRIMADIAG liquid-handling robot (EYOWN TECHNOLOGIES SL, Madrid, Spain). The ester was dispensed in replicates. After adding the esters, the 384-well plate was filled with 20 μL of 5 mM EPPS buffer, at pH 8.0, containing 0.912 mM Phenol Red (used as a pH indicator) using a QFill3 microplate filler. The final ester concentration of the ester in each well was 1.14 mg mL^{-1}, and the final concentration of Phenol Red was 0.45 mM. A total of 2 μL of protein extract (containing 1−5 mg mL^{-1} pure protein or 200 mg mL^{-1} wet cells expressing proteins) was immediately added to each well using an Eppendorf Repeater M4 pipet (Eppendorf, Hamburg, Germany) or a PRIMADIAG liquid-handling robot. Accordingly, the total reaction volume was 44 μL, with 4.5% (v/v) acetonitrile or DMSO in the reaction mixture. After incubation at 30 °C in a Synergy HT Multi-Mode Microplate Reader, ester hydrolysis was measured spectrophotometrically in continuous mode at 550 nm for a total time of 24 h. Commercially available CALA L and CALB L (Novozymes A/S, Bagsvaerd, Denmark) were diluted 10-fold with 5 mM EPPS buffer, at pH 8.0, and 2 μL of this solution was used immediately for reaction tests under the conditions described before. In all cases, specific activities (in U g^{-1} protein) were determined. One unit (U) of enzyme activity was defined as the amount of wet cells expressing EHs or pure EHs required to transform 1 μmol of substrate in 1 min under the assay conditions using the reported extinction coefficient ($\varepsilon_{\text{Phenol-red}}$ at 550 nm = 8450 M^{-1} cm^{-1}). All values were corrected for nonenzymatic transformation (i.e., the background rate) and for the background signal using *E. coli* cells that did not express any target protein (control cells included empty vectors). Note that a positive reaction was indicated by the restrictive criterion of a change greater than 6-fold above the background signal. Specific activity determinations (in U g^{-1}) for wet cells expressing each of the selected EHs or pure or commercial proteins are available in Supporting Information Tables S3 and S6, respectively.

Structural Determinations and Homology Modeling. The proteins EH1 and EH102 were expressed, purified, and crystallized using the sitting-drop method in Intelliplate 96-well plates and a Mosquito liquid-handling robot (TTP LabTech) according to previously described procedures.[37] For EHs for which crystal structures were not available, homology models were developed using Prime software from Schrodinger. Prime uses BLAST (with BLOSUM62 matrix) for homology search and alignment and refines the results using the Pfam database and pairwise alignment with ClustalW.

Protein Energy Landscape Exploration (PELE) Simulations. We used Protein Energy Landscape Exploration (PELE) software to sample the binding modes of glyceryl triacetate with EH1 and EH102.[38,39] The initial structures were taken from the coordinates of the EH1 and EH102 crystal structures (PDB codes: 5JD4 and 5JD3, respectively). The protonation state of titratable residues was estimated with the Protein Preparation Wizard (PROPKA)[40] and the H++ server (http://biophysics.cs.vt.edu/H++) followed by visible inspection. At pH 8 (the pH at which the activity assays were performed), the catalytic triad histidine residues were δ-protonated, and the catalytic triad aspartic acid residues were deprotonated, resulting in the formation of a histidine-serine and histidine-aspartic hydrogen-bonding network. The glyceryl triacetate structure was fully optimized with Jaguar[41] in an implicit solvent, and the electrostatic potential charges were computed with the density functional M06 at the 6-31G* level of theory. The ligand parameters were extracted from these for the classic simulations.

Cavity Volume and Solvent Accessible Surface Area (SASA) calculation. The relative Solvent Accessible Surface Area (SASA) for a residue was obtained using the GetArea Web server.[42] Cavity volumes were computed with Fpocket,[43] a very fast open-source protein pocket (cavity) detection algorithm based on Voronoi tessellation. Fpocket includes two other programs (dpocket and tpocket) that allow the extraction of pocket descriptors and the testing of owned scoring functions, respectively.

For extensive details of the methods, see Supporting Information Methods.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acschembio.7b00996.

Supporting Results, Methods, Figures S1−S7, and Table S4 (PDF)

Tables S1−S3, S5, and S6 (XLS)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: victor.guallar@bsc.es.

*E-mail: mferrer@icp.csic.es.

ORCID

Gerard Santiago: 0000-0002-0506-3049

Jurgen Pleiss: 0000-0003-1045-8202

Alexander F. Yakunin: 0000-0003-0813-6490

Víctor Guallar: 0000-0002-4580-1114

Manuel Ferrer: 0000-0003-4962-4714

Present Addresses

[γ]Current address: School of Chemistry, Bangor University, LL57 2UW Bangor, UK.

[δ]Current address: Lehrstuhl fur Biotechnologie, RWTH Aachen University, Aachen, Germany.

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

[e]Current address: Carl R. Woese Institute for Genomic Biology, Urbana, USA.

Author Contributions

[ζ]These authors contributed equally to this work.

Author Contributions

[ψ]These authors contributed equally in coordinating activities.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

C.C. thanks the Spanish Ministry of Economy, Industry and Competitiveness for a Ph.D. fellowship (Grant BES-2015-073829). V.M. thanks the Francisco Jose de Caldas Scholarship Program (Administrative Department of Science, Technology and Innovation, COLCIENCIAS). The authors acknowledge the members of the MAMBA, MAGICPAH, ULIXES, KILLSPILL and INMARE Consortia for their support in sample collection. David Rojo is also acknowledged for his valuable help with log P calculations. This project received funding from the European Union's Horizon 2020 research and innovation program [Blue Growth: Unlocking the potential of Seas and Oceans] under grant agreement no. 634486 (project acronym INMARE). This research was also supported by the European Community Projects MAGICPAH (FP7-KBBE-2009-245226), ULIXES (FP7-KBBE-2010-266473), and KILLSPILL (FP7-KBBE-2012-312139) and grants BIO2011-25012, PCIN-2014-107, BIO2014-54494-R, and CTQ2016-79138-R from the Spanish Ministry of Economy, Industry and Competitiveness. The present investigation was also funded by the Spanish Ministry of Economy, Industry and Competitiveness within the ERA NET IB2, grant no. ERA-IB-14-030 (MetaCat), the UK Biotechnology and Biological Sciences Research Council (BBSRC), grant no. BB/M029085/1, and the German Research Foundation (FOR1296). R.B. and P.N.G. acknowledge the support of the Supercomputing Wales project, which is part-funded by the European Regional Development Fund (ERDF) *via* the Welsh Government. O.V.G. and P.N.G. acknowledge the support of the Centre of Environmental Biotechnology Project funded by the European Regional Development Fund (ERDF) through the Welsh Government. A.Y. and A.S. gratefully acknowledge funding from Genome Canada (2009-OGI-ABC-1405) and the NSERC Strategic Network grant IBN. A.I.P. was supported by the Counseling of Economy and Employment of the Principality of Asturias, Spain (Grant FC-15-GRUPIN14-107). V.G. acknowledges the joint BSC-CRG-IRB Research Program in Computational Biology. The authors gratefully acknowledge financial support provided by the European Regional Development Fund (ERDF).

■ REFERENCES

(1) Huang, H., Pandya, C., Liu, C., Al-Obaidi, N. F., Wang, M., Zheng, L., Toews Keating, S., Aono, M., Love, J. D., Evans, B., Seidel, R. D., Hillerich, B. S., Garforth, S. J., Almo, S. C., Mariano, P. S., Dunaway-Mariano, D., Allen, K. N., and Farelli, J. D. (2015) Panoramic view of a superfamily of phosphatases through substrate profiling. *Proc. Natl. Acad. Sci. U. S. A. 112*, E1974−1983.

(2) Huang, R., Hippauf, F., Rohrbeck, D., Haustein, M., Wenke, K., Feike, J., Sorrelle, N., Piechulla, B., and Barkman, T. J. (2012) Enzyme functional evolution through improved catalysis of ancestrally non-preferred substrates. *Proc. Natl. Acad. Sci. U. S. A. 109*, 2966−2971.

(3) Yip, S. H., and Matsumura, I. (2013) Substrate ambiguous enzymes within the *Escherichia coli* proteome offer different evolutionary solutions to the same problem. *Mol. Biol. Evol. 30*, 2001−2012.

(4) Price, D. R., and Wilson, A. C. (2014) Substrate ambiguous enzyme facilitates genome reduction in an intracellular symbiont. *BMC Biol. 12*, 110.

(5) Devamani, T., Rauwerdink, A. M., Lunzer, M., Jones, B. J., Mooney, J. L., Tan, M. A., Zhang, Z. J., Xu, J. H., Dean, A. M., and Kazlauskas, R. J. (2016) Catalytic promiscuity of ancestral esterases and hydroxynitrile lyases. *J. Am. Chem. Soc. 138*, 1046−1056.

(6) Hult, K., and Berglund, P. (2007) Enzyme promiscuity: mechanism and applications. *Trends Biotechnol. 25*, 231−238.

(7) Copley, S. D. (2015) An evolutionary biochemist's perspective on promiscuity. *Trends Biochem. Sci. 40*, 72−78.

(8) London, N., Farelli, J. D., Brown, S. D., Liu, C., Huang, H., Korczynska, M., Al-Obaidi, N. F., Babbitt, P. C., Almo, S. C., Allen, K. N., and Shoichet, B. K. (2015) Covalent docking predicts substrates for haloalkanoate dehalogenase superfamily phosphatases. *Biochemistry 54*, 528−537.

(9) Nobeli, I., Favia, A. D., and Thornton, J. M. (2009) Protein promiscuity and its implications for biotechnology. *Nat. Biotechnol. 27*, 157−167.

(10) Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W. R., Golyshina, O. V., and Golyshin, P. N. (2016) Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microb. Biotechnol. 9*, 22−34.

(11) Braakman, R., and Smith, E. (2014) Metabolic evolution of a deep-branching hyperthermophilic chemoautotrophic bacterium. *PLoS One 9*, e87950.

(12) Giovannoni, S. J., Cameron Thrash, J., and Temperton, B. (2014) Implications of streamlining theory for microbial ecology. *ISME J. 8*, 1553−1565.

(13) Lan, T., Wang, X. R., and Zeng, Q. Y. (2013) Structural and functional evolution of positively selected sites in pine glutathione S-transferase enzyme family. *J. Biol. Chem. 288*, 24441−24451.

(14) Ferrer, M., Bargiela, R., Martínez-Martínez, M., Mir, J., Koch, R., Golyshina, O. V., and Golyshin, P. N. (2015) Biodiversity for biocatalysis: A review of the α/β-hydrolase fold superfamily of esterases-lipases discovered in metagenomes. *Biocatal. Biotransform. 33*, 235−249.

(15) Schmid, A., Dordick, J. S., Hauer, B., Kiener, A., Wubbolts, M., and Witholt, B. (2001) Industrial biocatalysis today and tomorrow. *Nature 409*, 258−268.

(16) Koudelakova, T., Chovancova, E., Brezovsky, J., Monincova, M., Fortova, A., Jarkovsky, J., and Damborsky, J. (2011) Substrate specificity of haloalkane dehalogenases. *Biochem. J. 435*, 345−354.

(17) Risso, V. A., Gavira, J. A., Mejia-Carmona, D. F., Gaucher, E. A., and Sanchez-Ruiz, J. M. (2013) Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian β-lactamases. *J. Am. Chem. Soc. 135*, 2899−2902.

(18) Amin, S. R., Erdin, S., Ward, R. M., Lua, R. C., and Lichtarge, O. (2013) Prediction and experimental validation of enzyme substrate specificity in protein structures. *Proc. Natl. Acad. Sci. U. S. A. 110*, E4195−4202.

(19) Anton, B. P., Chang, Y. C., Brown, P., Choi, H. P., Faller, L. L., Guleria, J., Hu, Z., Klitgord, N., Levy-Moonshine, A., Maksad, A., Mazumdar, V., McGettrick, M., Osmani, L., Pokrzywa, R., Rachlin, J., Swaminathan, R., Allen, B., Housman, G., Monahan, C., Rochussen, K., Tao, K., Bhagwat, A. S., Brenner, S. E., Columbus, L., de Crecy-Lagard, V., Ferguson, D., Fomenkov, A., Gadda, G., Morgan, R. D., Osterman, A. L., Rodionov, D. A., Rodionova, I. A., Rudd, K. E., Soll, D., Spain, J., Xu, S. Y., Bateman, A., Blumenthal, R. M., Bollinger, J. M., Chang, W. S., Ferrer, M., Friedberg, I., Galperin, M. Y., et al. (2013) The COMBREX project: design, methodology, and initial results. *PLoS Biol. 11*, e1001638.

(20) Colin, P. Y., Kintses, B., Gielen, F., Miton, C. M., Fischer, G., Mohamed, M. F., Hyvonen, M., Morgavi, D. P., Janssen, D. B., and Hollfelder, F. (2015) Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. *Nat. Commun. 6*, 10008.

(21) Chen, C., Huang, H., and Wu, C. H. (2017) Protein bioinformatics databases and resources. *Methods Mol. Biol. 1558*, 3−39.

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

(22) Fischer, M., and Pleiss, J. (2003) The Lipase Engineering Database: a navigation and analysis tool for protein families. *Nucleic Acids Res. 31*, 319−321.
(23) Arpigny, J. L., and Jaeger, K. E. (1999) Bacterial lipolytic enzymes: classification and properties. *Biochem. J. 343*, 177−183.
(24) Alcaide, M., Tornes, J., Stogios, P. J., Xu, X., Gertler, C., Di Leo, R., Bargiela, R., Lafraya, A., Guazzaroni, M. E., Lopez-Cortes, N., Chernikova, T. N., Golyshina, O. V., Nechitaylo, T. Y., Plumeier, I., Pieper, D. H., Yakimov, M. M., Savchenko, A., Golyshin, P. N., and Ferrer, M. (2013) Single residues dictate the co-evolution of dual esterases: MCP hydrolases from the α/β hydrolase family. *Biochem. J. 454*, 157−166.
(25) Lombard, V., Bernard, T., Rancurel, C., Brumer, H., Coutinho, P. M., and Henrissat, B. (2010) A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochem. J. 432*, 437−444.
(26) Popovic, A., Hai, T., Tchigvintsev, A., Hajighasemi, M., Nocek, B., Khusnutdinova, A. N., Brown, G., Glinos, J., Flick, R., Skarina, T., Chernikova, T. N., Yim, V., Bruls, T., Paslier, D. L., Yakimov, M. M., Joachimiak, A., Ferrer, M., Golyshina, O. V., Savchenko, A., Golyshin, P. N., and Yakunin, A. F. (2017) Activity screening of environmental metagenomic libraries reveals novel carboxylesterase families. *Sci. Rep. 7*, 44103.
(27) Janes, L. E., Lowendahl, C., and Kazlauskas, R. J. (1998) Rapid quantitative screening of hydrolases using pH indicators. Finding enantioselective hydrolases. *Chem. Eur. J. 4*, 2317−2324.
(28) Martínez-Martínez, M., Alcaide, M., Tchigvintsev, A., Reva, O., Polaina, J., Bargiela, R., Guazzaroni, M. E., Chicote, A., Canet, A., Valero, F., Rico Eguizabal, E., Guerrero, Mdel C., Yakunin, A. F., and Ferrer, M. (2013) Biochemical diversity of carboxyl esterases and lipases from Lake Arreo (Spain): a metagenomic approach. *Appl. Environ. Microbiol. 79*, 3553−3562.
(29) Daiha, K. d. G., Angeli, R., de Oliveira, S. D., and Almeida, R. V. (2015) Are lipases still important biocatalysts? A study of scientific publications and patents for technological forecasting. *PLoS One 10*, e0131624.
(30) Borrelli, K. W., Cossins, B., and Guallar, V. (2009) Exploring hierarchical refinement techniques for induced fit docking with protein and ligand flexibility. *J. Comput. Chem. 31*, 1224−1235.
(31) Hernandez-Ortega, A., Borrelli, K., Ferreira, P., Medina, M., Martínez, A. T., and Guallar, V. (2011) Substrate diffusion and oxidation in GMC oxidoreductases: an experimental and computational study on fungal aryl-alcohol oxidase. *Biochem. J. 436*, 341−350.
(32) Santiago, G., de Salas, F., Lucas, F., Monza, E., Acebes, S., Martínez, A., Camarero, S., and Guallar, V. (2016) Computer-aided laccase engineering: toward biological oxidation of arylamines. *ACS Catal. 6*, 5415−5423.
(33) Barak, Y., Nov, Y., Ackerley, D. F., and Matin, A. (2008) Enzyme improvement in the absence of structural knowledge: a novel statistical approach. *ISME J. 2*, 171−179.
(34) Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D. R., Alberti, A., Cornejo-Castillo, F. M., Costea, P. I., Cruaud, C., d'Ovidio, F., Engelen, S., Ferrera, I., Gasol, J. M., Guidi, L., Hildebrand, F., Kokoszka, F., Lepoivre, C., Lima-Mendez, G., Poulain, J., Poulos, B. T., Royo-Llonch, M., Sarmento, H., Vieira-Silva, S., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Bowler, C., de Vargas, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Jaillon, O., Not, F., Ogata, H., Pesant, S., Speich, S., Stemmann, L., Sullivan, M. B., Weissenbach, J., Wincker, P., Karsenti, E., Raes, J., Acinas, S. G., Bork, P., et al. (2015) Structure and function of the global ocean microbiome. *Science 348*, 1261359.
(35) Moult, J., Fidelis, K., Kryshtafovych, A., Schwede, T., and Tramontano, A. (2016) Critical assessment of methods of protein structure prediction: Progress and new directions in round XI. *Proteins: Struct., Funct., Genet. 84* (Suppl 1), 4−14.
(36) Barrozo, A., Duarte, F., Bauer, P., Carvalho, A. T. P., and Kamerlin, S. C. L. (2015) Cooperative electrostatic interactions drive functional evolution in the alkaline phosphatase superfamily. *J. Am. Chem. Soc. 137*, 9061−9076.
(37) Alcaide, M., Stogios, P. J., Lafraya, A., Tchigvintsev, A., Flick, R., Bargiela, R., Chernikova, T. N., Reva, O. N., Hai, T., Leggewie, C. C., Katzke, N., La Cono, V., Matesanz, R., Jebbar, M., Jaeger, K. E., Yakimov, M. M., Yakunin, A. F., Golyshin, P. N., Golyshina, O. V., Savchenko, A., and Ferrer, M. (2015) Pressure adaptation is linked to thermal adaptation in salt-saturated marine habitats. *Environ. Microbiol. 17*, 332−345.
(38) Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., and Jorgensen, W. L. (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B 105*, 6474−6487.
(39) Borrelli, K. W., Vitalis, A., Alcantara, R., and Guallar, V. (2005) PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *J. Chem. Theory Comput. 1*, 1304−1311.
(40) Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013) Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput.-Aided Mol. Des. 27*, 221−234.
(41) Bochevarov, A. D., Harder, E., Hughes, T. F., Greenwood, J. R., Braden, D. A., Philipp, D. M., Rinaldo, D., Halls, M. D., Zhang, J., and Friesner, R. A. (2013) Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *Int. J. Quantum Chem. 113*, 2110−2142.
(42) Fraczkiewicz, R., and Braun, W. (1998) Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem. 19*, 319.
(43) Le Guilloux, V., Schmidtke, P., and Tuffery, P. (2009) Fpocket: An open source platform for ligand pocket detection. *BMC Bioinf. 10*, 168.

DOI: 10.1021/acschembio.7b00996
ACS Chem. Biol. 2018, 13, 225−234

Chapter 2: Relationships between substrate promiscuity and chiral selectivity of esterases from phylogenetically and environmentally diverse microorganisms.

This contribution is an extension of the one described in Chapter 1. While the main objective of the previous Chapter was to identify substrate promiscuous ester-hydrolases and to establish the reasons why they have a broader specificity compared to others, in Chapter 2 we investigated the relationship between the substrate promiscuity and the enantio-selectivity and, in consequence, evaluated whether substrate promiscuous ester-hydrolases are also enantio-specific or not. For that, we have analyzed the biocatalytic data generated for the 145 ester-hydrolases described in Chapter 1. Particularly, we have focused on the analysis of the activity against 20 chiral compounds which were included in the 96 ester library used for substrate fingerprint. The apparent enantio-selectivity was measured by the ratio of the specific activity of the preferred over the non-preferred chiral ester when both chiral esters were tested separately, which is called the apparent selectivity factor.

We have found that promiscuous ester-hydrolases, that is, with broad substrate specificity, are commonly not enantio-specific. This could be a foreseeable result as it is known that promiscuous enzymes can accommodate a wide variety of substrate molecules into its active site. However, we were able to accurately determine the percentage of enantio-selective ester-hydrolases relative to the total of ester-hydrolases in our collection. As long as we know that most promiscuous ester-hydrolases are not enantio-selective in nature, we have been further working to introduce this capability (stringent stereo-specificity) artificially, as it will be shown Chapters 4 and 5 of this Doctoral thesis. For that, the enzyme with broader substrate spectra but being not enantio-specific was taken as a working model.

catalysts

MDPI

Communication

Relationships between Substrate Promiscuity and Chiral Selectivity of Esterases from Phylogenetically and Environmentally Diverse Microorganisms

Cristina Coscolín [1,†], Mónica Martínez-Martínez [1,†], Jennifer Chow [2], Rafael Bargiela [1,3], Antonio García-Moyano [4], Gro E. K. Bjerga [4], Alexander Bollinger [5], Runar Stokke [6], Ida H. Steen [6], Olga V. Golyshina [7,8], Michail M. Yakimov [9,10], Karl-Erich Jaeger [5,11], Alexander F. Yakunin [12], Wolfgang R. Streit [2], Peter N. Golyshin [7,8] and Manuel Ferrer [1,*]

1 Institute of Catalysis, Consejo Superior de Investigaciones Científicas, 28049 Madrid, Spain; cristina.coscolin@csic.es (C.C.); m.martinez@csic.es (M.M.-M.); rafaelbb@icp.csic.es (R.B.)
2 Biozentrum Klein Flottbek, Mikrobiologie & Biotechnologie, Universität Hamburg, 22609 Hamburg, Germany; jennifer.chow@uni-hamburg.de (J.C.); wolfgang.streit@uni-hamburg.de (W.R.S.)
3 School of Chemistry, Bangor University, Bangor LL57 2UW, UK; f.bargiela@bangor.ac.uk
4 Uni Research AS, Center for Applied Biotechnology, 5006 Bergen, Norway; antonio.Moyano@uib.no (A.G.-M.); Gro.Bjerga@uni.no (G.E.K.B.)
5 Institute of Molecular Enzyme Technology, Heinrich-Heine-University Düsseldorf, 52426 Jülich, Germany; a.bollinger@fz-juelich.de (A.B.); k.-e.jaeger@fz-juelich.de (K.-E.J.)
6 Department of Biology and KG Jebsen Centre for Deep Sea Research, University of Bergen, 5020 Bergen, Norway; Runar.Stokke@uib.no (R.S.); Ida.Steen@uib.no (I.H.S.)
7 School of Biological Sciences, Bangor University, LL57 2UW Bangor, UK; o.golyshina@bangor.ac.uk (O.V.G.); p.golyshin@bangor.ac.uk (P.N.G.)
8 Centre for Environmental Biotechnology, Bangor University, Bangor LL57 2UW, UK
9 Institute for Coastal Marine Environment, Consiglio Nazionale delle Ricerche, 98122 Messina, Italy; michail.yakimov@iamc.cnr.it
10 Immanuel Kant Baltic Federal University, 236040 Kaliningrad, Russia
11 Institute for Bio- and Geosciences IBG-1: Biotechnology, Forschungszentrum Jülich GmbH, 52426 Jülich, Germany
12 Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, ON M5S 3E5, Canada; a.iakounine@utoronto.ca
* Correspondence: mferrer@icp.csic.es; Tel.: +34-91-5854872
† These authors contributed equally to this work.

Received: 22 December 2017; Accepted: 3 January 2018; Published: 5 January 2018

Abstract: Substrate specificity and selectivity of a biocatalyst are determined by the protein sequence and structure of its active site. Finding versatile biocatalysts acting against multiple substrates while at the same time being chiral selective is of interest for the pharmaceutical and chemical industry. However, the relationships between these two properties in natural microbial enzymes remain underexplored. Here, we performed an experimental analysis of substrate promiscuity and chiral selectivity in a set of 145 purified esterases from phylogenetically and environmentally diverse microorganisms, which were assayed against 96 diverse esters, 20 of which were enantiomers. Our results revealed a negative correlation between substrate promiscuity and chiral selectivity in the evaluated enzymes. Esterases displaying prominent substrate promiscuity and large catalytic environments are characterized by low chiral selectivity, a feature that has limited commercial value. Although a low level of substrate promiscuity does not guarantee high chiral selectivity, the probability that esterases with smaller active sites possess chiral selectivity factors of interest for industry (>25) is significantly higher than for promiscuous enzymes. Together, the present study unambiguously demonstrates that promiscuous and selective esterases appear to be rare in nature and that substrate promiscuity can be used as an indicator of the chiral selectivity level of esterases, and vice versa.

Keywords: esterase; metagenomics; promiscuity; selectivity

1. Introduction

Presently, there is a great need for suitable biocatalysts with high process performance as greener alternatives to chemical synthesis [1,2]. It is expected that up to 40% of bulk chemical synthesis processes could be substituted by enzymatic catalysis by 2020 [1]. Along with requirements of a technical nature, such as process development and optimization, it is however widely recognized that the establishment of enzymatic processes is mainly a problem of finding, optimizing, or designing new and/or better performing enzymes. Nature is a rich reservoir from where enzymes can be isolated [3,4], because they are continuously evolving as a consequence of natural selection. Promiscuous enzymes are effective for converting multiple substrates, thus, they are industrially relevant [4–6]. Enzymes need to also be robust and, preferably, chiral selective to reduce raw material costs in the synthesis of pure chiral compounds [1,2,4,7]. That is, they need to be able to cleave preferentially only one chiral ester when offered a racemic mixture. Is it possible to find versatile enzymes displaying prominent substrate range and stringent chiral selectivity? Evaluating this possibility was the starting point of the present study.

In this study, we are interested in investigating as model enzymes serine ester hydrolases, hereafter referred to as esterases, from the structural superfamily of α/β-hydrolases. The activity of these esterases relies mainly on a catalytic triad usually formed by Ser, Asp/Glu, and His [8]. This enzyme class was selected for a number of reasons: it is widely distributed in the environment, it has important physiological functions, it includes hydrolases that are among the most important industrial biocatalysts, and extensive biochemical knowledge has been accumulated [4,5,7].

Just focusing on those from uncultivated microorganisms discovered through metagenomic approaches, esterases with prominent chiral selectivity have been identified and their use in the kinetic resolution of a number of esters is reported. Recent examples include those preferably hydrolyzing one of the chiral esters in racemic mixtures of ibuprofen esters [9,10]; ketoprofen esters [11–14]; solketal esters [15]; esters of phenylalkyl carboxylic acids, 1,1,1-trifluoro-2-phenylbut-3-yn-2-yl acetate and 3,7-dimethyl-1,6-octadien-3-yl acetate [16,17]; methyl 3-phenylglycidate [18]; 1-phenylethyl acetate [19,20]; ofloxacin butyl ester [21]; 1-octin-3-ol, 3-chlor-1-phenyl-1-propanol, trimethylsilylbutinol, *cis*/*trans*-1,2-cyclohexanediol, and isopropylidenglycerol acetate [22]; glycidyl butyrate [23]; methyl-mandelate, glycidyl-4-nitrobenzoate, methyl-3-bromo-2-methyl propionate, methyl lactate, menthyl acetate, neomenthyl acetate, pantolactone, and methyl 3-hydroxybutyrate [22,24,25]; 1-octin-3-ol, 3-chlor-1-phenyl-1-propanol, and trimethylsilylbutinol [22]; methyl-3-hydroxy-2-methylpropionate [26]; and esters of secondary alcohols [27,28], to cite some. The advances in metagenomics techniques and screening methods have allowed the discovery of these and other selective esterases [29]. These studies exemplify that esterases with selective character occur naturally, and that their chiral preference depend on structural factors in the proximity of the active-site. However, whether the selective character of these esterases, and many others, is linked to a broad or a narrow substrate spectrum has not been investigated, due to limited substrate sets employed.

Here, we investigate the relationships between the level of substrate promiscuity and chiral selectivity of a large set of 145 phylogenetically and environmentally diverse microbial esterases, whose specific activity against 96 distinct esters that included 20 chiral esters have been recently reported [5]. We provide unambiguous experimental evidence suggesting a negative association between substrate specificity and chiral selectivity in native esterases.

2. Results and Discussion

2.1. Relationships between Substrate Promiscuity and Chiral Selectivity

We have recently described an extensive analysis of the substrate spectra of 145 phylogenetically and environmentally diverse microbial esterases [5]. Experimental data on substrate conversion (i.e., units g^{-1} or $U \cdot g^{-1}$) followed for 24 h, at pH 8.0 and 30 °C was reported for 96 distinct esters. They included esters with variation in size of acyl and alcohol groups and with growing residues (aromatic, aliphatic, branched, and unbranched), halogenated esters, sugar esters, lactones, an alkyl

2.1. Relationships between Substrate Promiscuity and Chiral Selectivity

We have recently described an extensive analysis of the substrate spectra of 145 phylogenetically and environmentally diverse microbial esterases [5]. Experimental data on substrate conversion (i.e., units g^{-1} or $U \cdot g^{-1}$) followed for 24 h, at pH 8.0 and 30 °C was reported for 96 distinct esters. They included esters with variation in size of acyl and alcohol groups and with growing residues (aromatic, aliphatic, branched, and unbranched), halogenated esters, sugar esters, lactones, an alkyl di-ester, and 20 chiral esters (including (*R*) and (*S*) enantiomers of menthyl acetate, *N*-benzyl-proline ethyl ester, methyl mandelate, ethyl 4-chloro-3-hydroxybutyrate, methyl 3-hydroxybutyrate, methyl 3-hydroxyvalerate, neomenthyl acetate, methyl and ethyl lactate, and pantolactone). By meaning of the partitioning coefficient (log *p* value), which reflects electronic and steric effects and hydrophobic and hydrophilic characteristics, the 96 esters do show a broad chemical and structural variability [5]. This chemical variability also characterized the chiral esters tested (Figure 1).

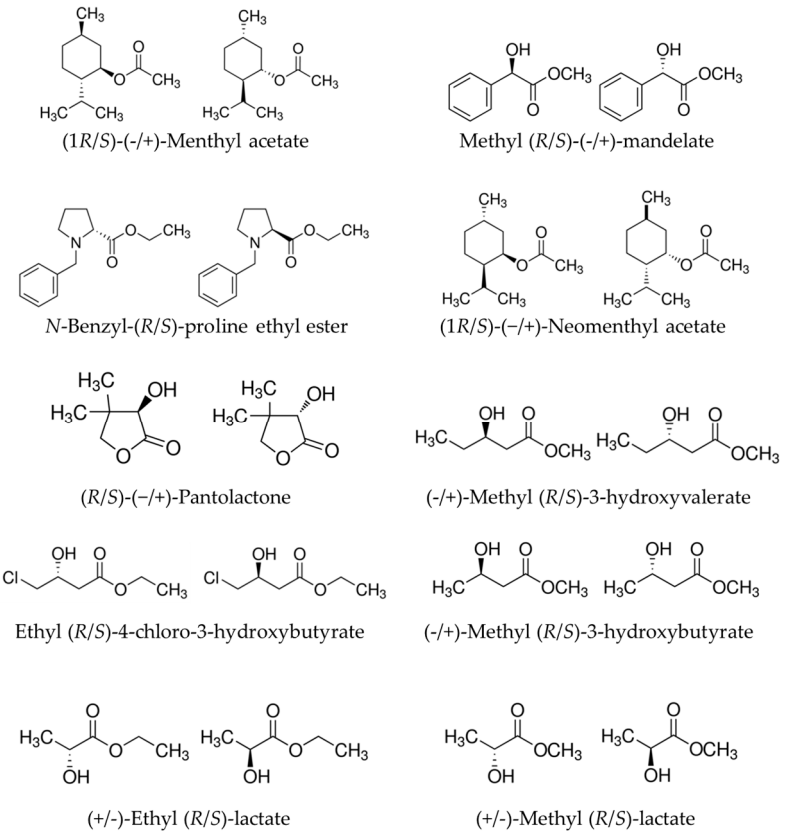


Figure 1. Representative chemical structures of 20 chiral esters used to evaluate chiral selectivity.

To find the relationships between substrate promiscuity and chiral selectivity we calculated the chiral selectivity factor for each of the 145 esterases and the 20 chiral esters tested (i.e., two enantiomers per pair) that were included in the 96-ester library (Figure 1). Selectivity factor was calculated as the ratio of specific activity ($U \cdot g^{-1}$) of the preferred over the non-preferred chiral ester when both esters were tested separately [30] (see Materials and Methods). These calculations were extracted from datasets reported previously [5]. It should be mentioned that these apparent values may not correspond to true selectivity or enantiomeric factors calculated when the enzyme is confronted to a racemic mixture, because the rates of hydrolysis of the enantiomers were measured separately [30]. Nevertheless, recent studies have clearly demonstrated that apparent and true selectivity values closely match each other [15]. These values were plotted against the number of esters hydrolyzed by each of the esterases (Figure 2), previously reported for each of the esterases [5].

From the 145 esterases, 40 did not show appreciable activity under assay conditions for any of the chiral esters tested. From those being active against at least one of the chiral esters (105 in total), 80 esterases were characterized by selectivity factors below a threshold of 25. Although esterases with stringent selectivity are preferred, it is commonly considered that enzymes with selectivity factor of 25 or above begin to have commercial value [31]. On the other hand, we found 25 chiral selective esterases, as judged by a selectivity factor above 25 (Figures 2 and 3). Ten of them showed stringent selectivity, that is they were capable of hydrolyzing only one of the enantiomer (Figure 3). Twelve of

From the 145 esterases, 40 did not show appreciable activity under assay conditions for any of the chiral esters tested. From those being active against at least one of the chiral esters (105 in total), 80 esterases were characterized by selectivity factors below a threshold of 25. Although esterases with stringent selectivity are preferred, it is commonly considered that enzymes with selectivity factor of 25 or above begin to have commercial value [31]. On the other hand, we found 25 chiral selective esterases, as judged by a selectivity factor above 25 (Figures 2 and 3). Ten of them showed stringent selectivity, that is they were capable of hydrolyzing only one of the enantiomer (Figure 3). Twelve of them were characterized by selectivity factors ranging from 25.9 to 59.3, and three did show prominent selectivity factors ranging from 219 to 686 (Figures 2 and 3).

As shown in Figure 2, we found a negative association between the level of substrate promiscuity, by meaning of the number of esters hydrolyzed, and the chiral selectivity factor. In further detail, according to criteria previously established [5] we considered an esterase 'specific' if it hydrolyses 9 esters or fewer, as 'moderate to highly promiscuous' if it hydrolyses between 10 and 42 esters, and as 'prominently promiscuous' if it hydrolyses 42 or more esters. None of the 25 hydrolases which showed a selectivity factor ≥25 were prominently promiscuous. Rather, they were capable of accepting 36 or fewer substrates. However, not all hydrolases converting 36 or fewer esters were selective according to the 25-selective factor threshold. Indeed, only 25 out of 85 in total (or 29%) were selective, with different selectivity factors and chiral preferences (Figure 3). This is most likely due to the fact that the ability to selectively hydrolyze an enantiomer in a racemate may depend on the topology of the catalytic environment [5].

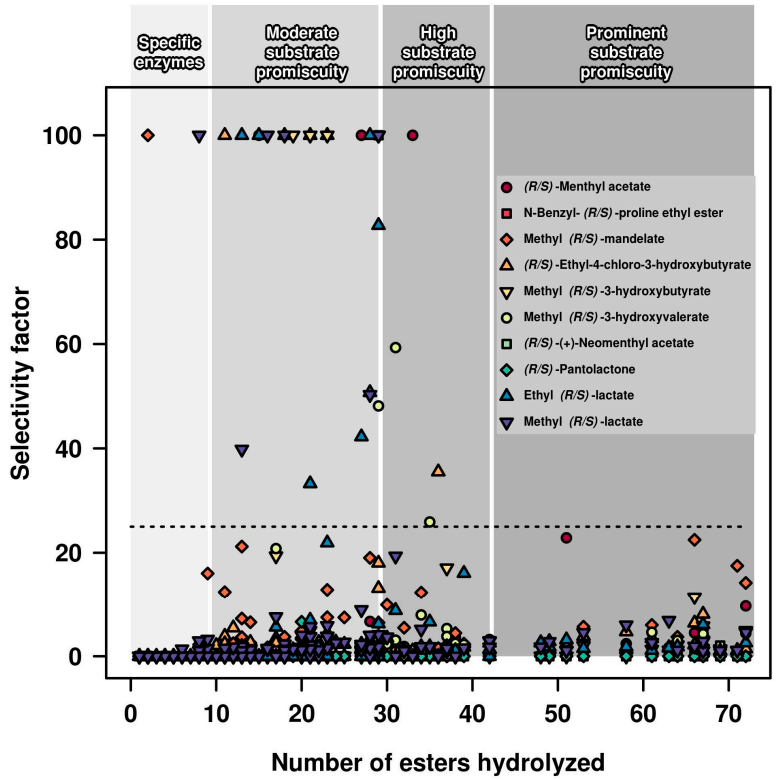


Figure 2. Chiral selectivity factor vs. number of esters hydrolyzed per each of the 145 hydrolases tested. Selectivity factor was calculated per each pair of enantiomers as the ratio of specific activity ($U{\cdot}g^{-1}$) of the preferred over the non-preferred chiral ester when each of the chiral esters was tested separately. Chiral esters are color coded. The value 100 was arbitrarily given to represent those esterases capable of hydrolyzing, under our assay conditions, only one of the enantiomers (100% selective) and those with selectivity factors higher than 100. These data are based on the data reported previously [5], using conditions described in Materials and Methods. The level of promiscuity, according to criteria previously established [5], is marked under a shadowed grey background. The 25-selectivity factor threshold at which an esterase started to have commercial value is indicated by a horizontal dashed gray line.

background. The 25-selectivity factor threshold at which an esterase started to have commercial value is indicated by a horizontal dashed gray line.

Catalysts **2018**, *8*, 10 5 of 8

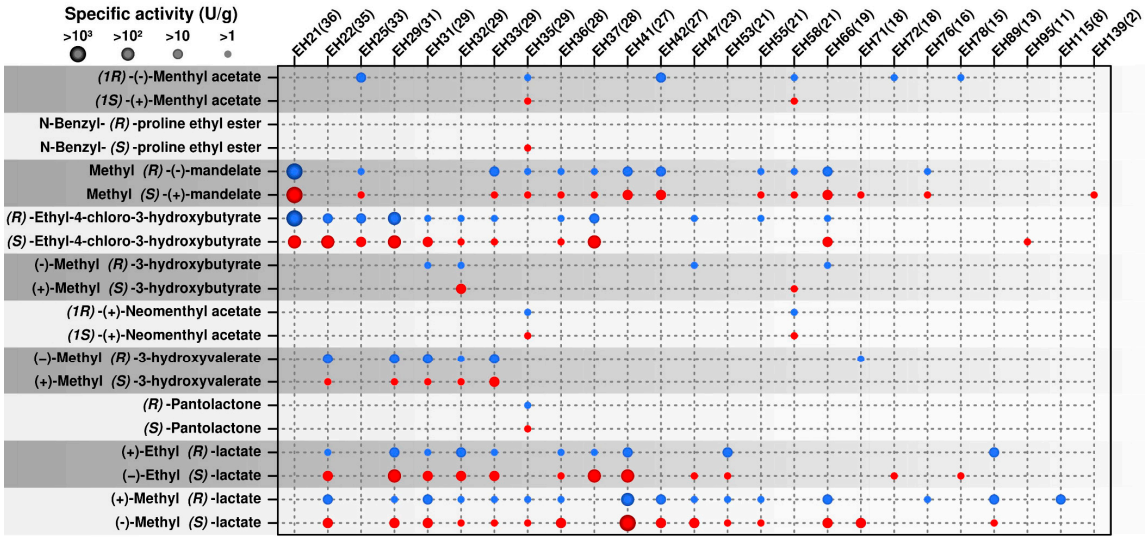


Figure 3. Chiral preferences of 25 hydrolases which were found to be selective for at least one chiral ester according to the 25-selective factor threshold. The figure illustrates the specific activity ($U \cdot g^{-1}$; represented by the size of the circles) of each esterase per each of the 20 chiral esters tested. The ID code for each esterase (for full description see ref. [5]) is shown on the top; the number of esters (out of 96 tested) hydrolyzed by each esterase is shown in brackets. The Figure was created with the R language console from data previously reported [5]. The list of the 20 chiral esters tested is shown on the left, with (*R*)-enantiomer in blue and (*S*)-enantiomer in red color. The protocol established and used to identify the esters hydrolyzed by each esterase is described in Materials and Methods.

2.2. Occurrence of Multi-Selective Esterases

Figure 3 summarizes the chiral preference of esterases fitting to the 25-selective factor threshold for each of the 20 chiral esters tested. As can be seen in Figure 3, esterases showed different level of promiscuity and preference for (*R*) or (*S*) methyl acetate, menthyl mandelate, methyl 3-hydroxybutyrate, *N*-benzyl-proline ethyl ester, ethyl-4-chloro-3-hydroxybutyrate, and (*R*)-ethyl lactate. Esterases selective for (*R*) or (*S*)-pantolactone, *N*-benzyl-proline ethyl ester, and neomenthyl acetate were the least abundant, suggesting these chiral esters are less preferred substrates. As shown in Figure 3, we also found that 5 out of 25 esterases fitting to the 25-selective threshold did show stringent selectivity or selectivity factor higher than 25 for several chiral esters differing in chemical and structural nature. They included one being selective for methyl 3-hydroxybutyrate (*R*-selective) and ethyl lactate (*S*-preference) (EH31), one for ethyl-4-chloro-3-hydroxybutyrate (*S*-selective) and methyl 3-hydroxybutyrate (*R*-selective) (EH47), one for ethyl lactate (*S*-selective) (EH71), one for methyl lactate (*S*-selective) and methyl 3-hydroxybutyrate (*S*-selective) (EH72), and one for ethyl lactate (*R*-selective) (EH78). The other 20 esterases did show chiral preference but with lower selectivity factors (Figure 3). This suggests that may have lower abundancy.

3. Materials and Methods

3.1. Source of Chemicals, Enzymes, and Datasets

All chemicals for which activity data are reported were of the purest grade available and were purchased as reported [5]. The present study used datasets of hydrolytic activity ($U \cdot g^{-1}$) for 145 esterases assayed at 550 nm using 96-structurally diverse esters in 384-well plates. Reactions were followed for 24 h, at pH 8.0 and 30 °C. Datasets are available elsewhere [5].

3.2. Selectivity Factor Calculation

The chiral selectivity factor is defined as the ratio of the specific activity [30] for each enantiomer, measured separately as described previously [5]. Briefly, reaction mixture contains 5 mM *N*-(2-hydroxyethyl)piperazine-*N*′-(3-propanesulfonic acid buffer, pH 8.0, 4.5% (*v*/*v*) acetonitrile or dimethyl sulfoxide, 0.45 mM Phenol Red (used as a pH indicator), a concentration of each of the esters of 1.14 mg·mL^{-1}, and 2 µg of proteins. Reactions were allowed to proceed kinetically at 30 °C and hydrolytic activity (U·g^{-1}) calculated followed for 24 h [5]. Selectivity factor was calculated considering the preferred over the non-preferred chiral ester, whatever the preferred (*R*) or (*S*) ester.

4. Conclusions

Herein, we show the value of the systematic investigation of enzyme activity to deepen our understanding of the relationships between substrate promiscuity and chiral selectivity. By comparing the number of esters that 145 diverse esterases hydrolyze as an indicator of the substrate promiscuity level and their selectivity factors as an indicator of enantio-selectivity, we found unambiguous evidence that esterases with broad substrate spectra do commonly show low selectivity for chiral molecules. In this study, the proportion of esterases with both prominent promiscuity and selectivity approaches zero percent. By contrast, the proportion of esterases with low to moderate promiscuity but prominent selectivity was as high as 29%. This suggests that the substrate promiscuity may be used as an indicator of the selective character of esterases. Promiscuous esterases acting against multiple substrates, while at the same time being enantio-selective, appear to be rare in nature, or at least in the habitats from where the esterases herein described were isolated [5]. As these enzymes are of interest for application purposes [1–6,32], protein engineering and rational design may be needed to obtain esterases being promiscuous and selective for industrial applications. We anticipate that the possibility to transform a promiscuous but not selective esterase into an efficient enantio-selective biocatalyst would require less engineering effort because increasing the selectivity for an enantiomer may involve a reduced number of contacts close to the active sites (for a recent example see reference [33]). Conversely, increasing the substrate spectra of a selective non-promiscuous esterase would require large rearrangement of the catalytic environment which may, at the same time, result in significant reduction or even loss of enantio-selectivity. This is because non-promiscuous esterases are characterized by catalytic environments that are highly exposed and have small volumes, while an esterase for being promiscuous requires a large active site volume and lower relative solvent accessible surface area [5], that are difficult to be designed through few mutations.

Acknowledgments: This project has received funding from the European Union's Horizon 2020 research and innovation program (Blue Growth: Unlocking the potential of Seas and Oceans) through the Project 'INMARE' under grant agreement No. 634486 and ERA-IB 5 'METACAT'. This work was further funded by grants PCIN-2014-107 (within ERA NET IB2 grant nr. ERA-IB-14-030—MetaCat), PCIN-2017-078 (within the ERA-MarineBiotech grant ProBone), BIO2014-54494-R and BIO2017-85522-R from the Spanish Ministry of Economy and Competitiveness. The present investigation was also funded by the UK Biotechnology and Biological Sciences Research Council (BBSRC), grant nr. BB/M029085/1. P.N.G. acknowledges the support of the Supercomputing Wales project, which is partly funded by the European Regional Development Fund (ERDF) via the Welsh Government. O.V.G. and P.N.G. acknowledge the support of the Centre of Environmental Biotechnology Project funded by the European Regional Development Fund (ERDF) through the Welsh Government. A.F.Y. gratefully acknowledges funding from Genome Canada (2009-OGI-ABC-1405) and the NSERC Strategic Network grant IBN. The authors gratefully acknowledge financial support provided by the European Regional Development Fund (ERDF). C. Coscolín thanks the Spanish Ministry of Economy, Industry and Competitiveness for a PhD fellowship (Grant BES-2015-073829).

Author Contributions: M.F., P.N.G., and W.R.S. conceived the study; C.C., M.M.-M., J.C., R.B., A.G.-M., G.E.K.B., A.B., R.S., I.H.S., O.V.G., M.M.Y., K.-E.J. and A.F.Y. contributed to enzyme collection and data analysis; M.F. drafted the manuscript which was revised by all co-authors.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Martínez-Martínez, M.; Bargiela, R.; Ferrer, M. Metagenomics and the search for industrial enzymes. In *Biotechnology of Microbial Enzymes*, 1st ed.; Brahmachari, G., Demain, A.L., Adrio, J.L., Eds.; Academic Press: Chennai, India, 2015; pp. 167–184.
2. Martínez-Martínez, M.; Bargiela, R.; Coscolín, C.; Navarro-Fernández, J.; Golyshin, P.N.; Ferrer, M. Functionalization and modification of hydrocarbon-like molecules guided by metagenomics: Enzymes most requested at the industrial scale for chemical synthesis as study cases. In *Consequences of Microbial Interactions with Hydrocarbons, Oils, and Lipids: Production of Fuels and Chemicals*; Lee, S.Y., Ed.; Springer International Publishing AG: Cham, Switzerland, 2016; pp. 1–26.
3. Yarza, P.; Yilmaz, P.; Pruesse, E.; Glöckner, F.O.; Ludwig, W.; Schleifer, K.H.; Whitman, W.B.; Euzéby, J.; Amann, R.; Rosselló-Móra, R. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* **2014**, *12*, 635–645. [CrossRef] [PubMed]
4. Ferrer, M.; Martínez-Martínez, M.; Bargiela, R.; Streit, W.R.; Golyshina, O.V.; Golyshin, P.N. Estimating the success of enzyme bioprospecting through metagenomics: Current status and future trends. *Microb. Biotechnol.* **2016**, *9*, 22–34. [CrossRef] [PubMed]
5. Martínez-Martínez, M.; Coscolín, C.; Santiago, G.; Chow, J.; Stogios, P.; Bargiela, R.; Gertler, C.; Navarro-Fernández, J.; Bollinger, A.; Thies, S.; et al. Determinants and prediction of esterase substrate promiscuity patterns. *ACS Chem. Biol.* **2017**. [CrossRef] [PubMed]
6. Schmid, A.; Dordick, J.S.; Hauer, B.; Kiener, A.; Wubbolts, M.; Witholt, B. Industrial biocatalysis today and tomorrow. *Nature* **2001**, *409*, 258–268. [CrossRef] [PubMed]
7. Ferrer, M.; Bargiela, R.; Martínez-Martínez, M.; Mir, J.; Koch, R.; Golyshina, O.V.; Golyshin, P.N. Biodiversity for Biocatalysis: A review for α/β-hydrolases of the esterase-lipase superfamily as case of Study. *Biocatal. Biotransform.* **2015**, *33*, 235–249. [CrossRef]
8. Aranda, J.; Cerqueira, N.M.; Fernandes, P.A.; Roca, M.; Tuñon, I.; Ramos, M.J. The Catalytic Mechanism of Carboxylesterases: A Computational Study. *Biochemistry* **2014**, *53*, 5820–5829. [PubMed]
9. Elend, C.; Schmeisser, C.; Hoebenreich, H.; Steele, H.L.; Streit, W.R. Isolation and characterization of a metagenome-derived and cold-active lipase with high stereospecificity for (*R*)-ibuprofen esters. *J. Biotechnol.* **2007**, *130*, 370–377. [CrossRef] [PubMed]
10. Chow, J.; Kovacic, F.; Dall Antonia, Y.; Krauss, U.; Fersini, F.; Schmeisser, C.; Lauinger, B.; Bongen, P.; Pietruszka, J.; Schmidt, M.; et al. The Metagenome-derived enzymes LipS and LipT increase the diversity of known lipases. *PLoS ONE* **2012**, *7*, e47665. [CrossRef] [PubMed]
11. Kim, Y.J.; Choi, G.S.; Kim, S.B.; Yoon, G.S.; Kim, Y.S.; Ryu, Y.W. Screening and characterization of a novel esterase from a metagenomic library. *Protein Expr. Purif.* **2006**, *45*, 315–323. [CrossRef] [PubMed]
12. Yoon, S.; Kim, S.; Ryu, Y.; Kim, T.D. Identification and characterization of a novel (*S*)-ketoprofen-specific esterase. *Int. J. Biol. Macromol.* **2007**, *41*, 1–7. [CrossRef] [PubMed]
13. Ngo, T.D.; Ryu, B.H.; Ju, H.; Jang, E.J.; Kim, K.K.; Kim, T.D. Crystallographic analysis and biochemical applications of a novel penicillin-binding protein/beta-lactamase homologue from a metagenomic library. *Acta Crystallogr. D Biol. Crystallogr.* **2014**, *70*, 2455–2466. [CrossRef] [PubMed]
14. Kim, J.; Seok, S.H.; Hong, E.; Yoo, T.H.; Seo, M.D.; Ryu, Y. Crystal structure and characterization of esterase Est25 mutants reveal improved enantioselectivity toward (*S*)-ketoprofen ethyl ester. *Appl. Microbiol. Biotechnol.* **2017**, *101*, 2333–2342. [CrossRef] [PubMed]
15. Ferrer, M.; Golyshina, O.V.; Chernikova, T.N.; Khachane, A.N.; Martins dos Santos, V.A.P.; Yakimov, M.M.; Timmis, K.N.; Golyshin, P.N. Microbial enzymes mined from the Urania deep-sea hypersaline anoxic basin. *Chem. Biol.* **2005**, *12*, 895–904. [CrossRef] [PubMed]
16. Kourist, R.; Hari Krishna, S.; Patel, J.S.; Bartnek, F.; Hitchman, T.S.; Weiner, D.P.; Bornscheuer, U.T. Identification of a metagenome-derived esterase with high enantioselectivity in the kinetic resolution of arylaliphatic tertiary alcohols. *Org. Biomol. Chem.* **2007**, *5*, 3310–3313. [CrossRef] [PubMed]
17. Fernández-Álvaro, E.; Kourist, R.; Winter, J.; Böttcher, D.; Liebeton, K.; Naumer, C.; Eck, J.; Leggewie, C.; Jaeger, K.E.; Streit, W.; et al. Enantioselective kinetic resolution of phenylalkyl carboxylic acids using metagenome-derived esterases. *Microb. Biotechnol.* **2010**, *3*, 59–64. [CrossRef] [PubMed]
18. Ouyang, L.M.; Liu, J.Y.; Qiao, M.; Xu, J.H. Isolation and biochemical characterization of two novel metagenome-derived esterases. *Appl. Biochem. Biotechnol.* **2013**, *169*, 15–28. [CrossRef] [PubMed]

19. Martini, V.; Glogauer, A.; Muller-Santos, M.; Lulek, J.; de Souza, E.; Mitchell, D.; Pedrosa, F.; Krieger, N. First co-expression of a lipase and its specific foldase obtained by metagenomics. *Microb. Cell Factories* **2014**, *13*, 171. [CrossRef] [PubMed]
20. Alnoch, R.C.; Martini, V.P.; Glogauer, A.; Costa, A.C.; Piovan, L.; Muller-Santos, M.; de Souza, E.M.; de Oliveira Pedrosa, F.; Mitchell, D.A.; Krieger, N. Immobilization and characterization of a new regioselective and enantioselective lipase obtained from a metagenomic library. *PLoS ONE* **2015**, *10*, e0114945. [CrossRef] [PubMed]
21. Jeon, J.H.; Kim, J.T.; Kang, S.G.; Lee, J.H.; Kim, S.J. Characterization and its potential application of two esterases derived from the arctic sediment metagenome. *Mar. Biotechnol.* **2009**, *11*, 307–316. [CrossRef] [PubMed]
22. Elend, C.; Schmeisser, C.; Leggewie, C.; Babiak, P.; Carballeira, J.D.; Steele, H.L.; Reymond, J.L.; Jaeger, K.E.; Streit, W.R. Isolation and biochemical characterization of two novel metagenome-derived esterases. *Appl. Environ. Microbiol.* **2006**, *72*, 3637–3645. [CrossRef] [PubMed]
23. Martínez-Martínez, M.; Alcaide, M.; Tchigvintsev, A.; Reva, O.; Polaina, J.; Bargiela, R.; Guazzaroni, M.-E.; Chicote, Á.; Canet, A.; Valero, F.; et al. Biochemical diversity of carboxyl esterases and lipases from Lake Arreo (Spain): A metagenomic approach. *Appl. Environ. Microbiol.* **2013**, *79*, 3553–3562. [CrossRef] [PubMed]
24. Alcaide, M.; Tchigvintsev, A.; Martinez-Martinez, M.; Popovic, A.; Reva, O.N.; Lafraya, A.; Bargiela, R.; Nechitaylo, T.Y.; Matesanz, R.; Cambon-Bonavita, M.A.; et al. Identification and characterization of carboxyl esterases of gill chamber-associated microbiota in the deep-sea shrimp *Rimicaris exoculata* by using functional metagenomics. *Appl. Environ. Microbiol.* **2015**, *81*, 2125–2136. [CrossRef] [PubMed]
25. Placido, A.; Hai, T.; Ferrer, M.; Chernikova, T.N.; Distaso, M.; Armstrong, D.; Yakunin, A.F.; Toshchakov, S.V.; Yakimov, M.M.; Kublanov, I.V.; et al. Diversity of hydrolases from hydrothermal vent sediments of the Levante Bay.; Vulcano Island (*Aeolian archipelago*) identified by activity-based metagenomics and biochemical characterization of new esterases and an arabinopyranosidase. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 10031–10046. [CrossRef] [PubMed]
26. Lee, H.W.; Jung, W.K.; Kim, Y.H.; Ryu, B.H.; Kim, T.D.; Kim, J.; Kim, H. Characterization of a novel alkaline family VIII esterase with *S*-enantiomer preference from a compost metagenomic library. *J. Microbiol. Biotechnol.* **2016**, *26*, 315–325. [CrossRef] [PubMed]
27. Gao, W.; Fan, H.; Chen, L.; Wang, H.; Wei, D. Efficient kinetic resolution of secondary alcohols using an organic solvent-tolerant esterase in non-aqueous medium. *Biotechnol. Lett.* **2016**, *38*, 1165–1171. [CrossRef] [PubMed]
28. Kumar, R.; Banoth, L.; Banerjee, U.C.; Kaur, J. Enantiomeric separation of pharmaceutically important drug intermediates using a metagenomic lipase and optimization of its large scale production. *Int. J. Biol. Macromol.* **2017**, *95*, 995–1003. [CrossRef] [PubMed]
29. Böttcher, D.; Zägel, P.; Schmidt, M.; Bornscheuer, U.T. A microtiter plate-based assay to screen for active and stereoselective hydrolytic enzymes in enzyme libraries. *Methods Mol. Biol.* **2017**, *1539*, 197–204. [PubMed]
30. Gawley, R.E. Do the terms "% ee" and "% de" make sense as expressions of stereoisomer composition or stereoselectivity? *J. Org. Chem.* **2006**, *71*, 2411–2416. [CrossRef] [PubMed]
31. Reetz, M.T. (Ed.) Introduction to directed evolution. In *Directed Evolution of Selective Enzymes: Catalysts for Organic Chemistry and Biotechnology*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2016; pp. 1–16.
32. Romano, D.; Bonomi, F.; de Mattos, M.C.; de Sousa Fonseca, T.; de Oliveira Mda, C.; Molinari, F. Esterases as stereoselective biocatalysts. *Biotechnol. Adv.* **2015**, *33*, 547–565. [CrossRef] [PubMed]
33. Wikmark, Y.; Svedendahl Humble, M.; Bäckvall, J.E. Combinatorial library based engineering of *Candida antarctica* lipase A for enantioselective transacylation of sec-alcohols in organic solvent. *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 4284–4288. [CrossRef] [PubMed]

Chapter 3: Bioprospecting reveals class III ω-transaminases converting bulky ketones and environmentally relevant polyamines.

The results presented in Chapter 1 demonstrated that establishing and characterizing a large collection of novel ester-hydrolases may allow deepening into the determinants of substrate promiscuity while at the same time identifying those enzymes performing better than commercial prototypes. Consequently, in this Chapter, we have extended this analysis to transaminases. Transaminases are very appreciated in biotechnological purposes, such as those for producing optically pure amines relevant in drug development. Through this Chapter, we recovered, by applying genomic and metagenomic tools, a set of ten class III ω-transaminases to investigate the molecular reasons for promiscuity and enantio-selectivity in this class of enzymes.

The reasoning of this investigation is the following. Most of the ω-transaminases, including those of the class III, described to date in the literature are S-stereospecific and they show little activity against bulky aldehydes, ketones and amines, which are among the most appreciable substrates in the industry. Nevertheless, by applying a novel functional screen method and by using a library of more than 36 different substrates, in this study we were able to recover a set of class III ω-transaminases with high capacity to convert bulky ketones and aldehydes. Among them, a sub-set of 4 enzymes showed the unusual capacity to convert (R)-amines, while the other 6 showed the common capacity to convert (S)-amines. Other biochemical characteristics were also experimentally calculated as optimal temperatures for activity, finding a range from 45 to 65°C. Additionally, few of our enzymes can retain 50% of its activity in up to 50% [vol/vol] in water-soluble solvents.

Using sequence alignment and 3D-modeling software programs we were able to find the reasons for the unusual properties of the class III ω-transaminases herein investigated compared to those previously described in the bibliography. We found that the presence of a hairpin region proximal to a highly conserved arginine residue, also called flipping arginine, is responsible for the capacity of such enzymes to convert or not bulky ketones and bulky aldehydes. In addition to that, the orientation of this arginine, which may be influenced by the presence of this hairpin region, is relevant for the differences in the enantio-selectivity. An outward position favors the capacity to convert (R)-amines, whereas the inward position that for (S)-amines. We additionally found that this configuration may have an *in vivo* role in the catabolism of putrescine, a naturally produced amine.

As we have identified two structural determinants defining the capacity to convert bulky substrates and (R)-amines, one can now and in the future engineer a class III ω-transaminases capable of converting bulky ketones and bulky aldehydes while carrying stringent (R)-selectivity for amines.

This work and that reported in Chapter 1, demonstrate how important is to access a high enzymatic diversity and establishing and working with a broad library of different substrates. This fills the gaps of information between the sequences and the catalytic profile associated with the enzymes encoded by those, and in combination with computational tools, for finding their structural determinants.

AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

BIOTECHNOLOGY

Bioprospecting Reveals Class III ω-Transaminases Converting Bulky Ketones and Environmentally Relevant Polyamines

Cristina Coscolín,[a] Nadine Katzke,[b] Antonio García-Moyano,[c] José Navarro-Fernández,[a] David Almendral,[a] Mónica Martínez-Martínez,[a] Alexander Bollinger,[b] Rafael Bargiela,[d] Christoph Gertler,[d]* Tatyana N. Chernikova,[d] David Rojo,[e] Coral Barbas,[e] Hai Tran,[d] Olga V. Golyshina,[d,f] Rainhard Koch,[g] Michail M. Yakimov,[h,i] Gro E. K. Bjerga,[c] Peter N. Golyshin,[d,f] Karl-Erich Jaeger,[b] Manuel Ferrer,[a] The INMARE Consortium

[a]Institute of Catalysis, Consejo Superior de Investigaciones Científicas, Madrid, Spain
[b]Institute of Molecular Enzyme Technology, Heinrich Heine University Düsseldorf and Forschungszentrum Jülich GmbH, Jülich, Germany
[c]NORCE Norwegian Research Centre AS, Bergen, Norway
[d]School of Natural Sciences, Bangor University, Bangor, United Kingdom
[e]Centro de Metabolómica y Bioanálisis (CEMBIO), Facultad de Farmacia, Universidad CEU San Pablo, Boadilla del Monte, Madrid, Spain
[f]Centre for Environmental Biotechnology, Bangor University, Bangor, United Kingdom
[g]Bayer AG, Engineering and Technology Department, Leverkusen, Germany
[h]Institute for Biological Resources and Marine Biotechnology (IRBIM-CNR), Messina, Italy
[i]Immanuel Kant Baltic Federal University, Kaliningrad, Russia

ABSTRACT Amination of bulky ketones, particularly in (*R*) configuration, is an attractive chemical conversion; however, known ω-transaminases (ω-TAs) show insufficient levels of performance. By applying two screening methods, we discovered 10 amine transaminases from the class III ω-TA family that were 38% to 76% identical to homologues. We present examples of such enzymes preferring bulky ketones over keto acids and aldehydes with stringent (*S*) selectivity. We also report representatives from the class III ω-TAs capable of converting (*R*) and (*S*) amines and bulky ketones and one that can convert amines with longer alkyl substituents. The preference for bulky ketones was associated with the presence of a hairpin region proximal to the conserved Arg414 and residues conforming and close to it. The outward orientation of Arg414 additionally favored the conversion of (*R*) amines. This configuration was also found to favor the utilization of putrescine as an amine donor, so that class III ω-TAs with Arg414 in outward orientation may participate *in vivo* in the catabolism of putrescine. The positioning of the conserved Ser231 also contributes to the preference for amines with longer alkyl substituents. Optimal temperatures for activity ranged from 45 to 65°C, and a few enzymes retained ≥50% of their activity in water-soluble solvents (up to 50% [vol/vol]). Hence, our results will pave the way to design, in the future, new class III ω-TAs converting bulky ketones and (*R*) amines for the production of high-value products and to screen for those converting putrescine.

IMPORTANCE Amine transaminases of the class III ω-TAs are key enzymes for modification of chemical building blocks, but finding those capable of converting bulky ketones and (*R*) amines is still challenging. Here, by an extensive analysis of the substrate spectra of 10 class III ω-TAs, we identified a number of residues playing a role in determining the access and positioning of bulky ketones, bulky amines, and (*R*)- and (*S*) amines, as well as of environmentally relevant polyamines, particularly putrescine. The results presented can significantly expand future opportunities for designing (*R*)-specific class III ω-TAs to convert valuable bulky ketones and amines, as well as for deepening the knowledge into the polyamine catabolic pathways.

KEYWORDS amine transaminases, biodiversity, chiral amine, metagenomics, putrescine, transaminase

Citation Coscolín C, Katzke N, García-Moyano A, Navarro-Fernández J, Almendral D, Martínez-Martínez M, Bollinger A, Bargiela R, Gertler C, Chernikova TN, Rojo D, Barbas C, Tran H, Golyshina OV, Koch R, Yakimov MM, Bjerga GEK, Golyshin PN, Jaeger K-E, Ferrer M, The INMARE Consortium. 2019. Bioprospecting reveals class III ω-transaminases converting bulky ketones and environmentally relevant polyamines. Appl Environ Microbiol 85:e02404-18. https://doi.org/10.1128/AEM.02404-18.

Editor Ning-Yi Zhou, Shanghai Jiao Tong University

Address correspondence to Manuel Ferrer, mferrer@icp.csic.es.

* Present address: Christoph Gertler, Lehrstuhl für Biotechnologie, RWTH Aachen University, Aachen, Germany.

C.C., N.K., and A.G.-M. contributed equally to this work. P.N.G., K.-E.J., and M.F. equally coordinated and conceived this work.

Received 6 October 2018
Accepted 4 November 2018
Accepted manuscript posted online 9 November 2018
Published 9 January 2019

Transaminases (TAs) (EC 2.6.1.x), also called aminotransferases, are versatile enzymes with industrial potential (1). They catalyze asymmetric amine transfer reactions between an amine and a ketone, aldehyde, or keto-acid and, thus, are key enzymes to produce building blocks for drug discovery and chemical biology. All transaminases reported so far require pyridoxal-5′-phosphate (PLP) as a coenzyme, which serves as a molecular shuttle for ammonia and electrons between the amine donor and the acceptor in a catalytic cycle. First, the amine group from the amine donor binds to the enzyme, and then pyridoxamine-5′-phosphate (PMP) is formed from PLP and the amine donor is released as a keto product. Afterwards, PMP transfers the amine group to the acceptor and PMP is regenerated to PLP, closing the catalytic cycle. Based on their amino acid sequences, transaminases are classified in six groups (classes I to VI) (1), with class III covering the so-called ω-transaminases (ω-TAs). Within the ω-TAs, the class of amine transaminases (ATAs) has industrial relevance, as they have been used for the preparation of optically pure amines starting from the corresponding ketones (1, 2).

In an ideal scenario, functional screening with genomics and metagenomics techniques would allow the identification of a new generation of microbial biocatalysts, including ATAs of the class III ω-TAs (3–6). However, extensive bioprospecting by metagenomics was only rarely successful (5), despite the growing number of sequences available in public databases (7). Indeed, only three class III ω-TAs have been identified by metagenomics techniques; however, this was by applying sequence homology-based techniques rather than functional methods (8). These enzymes showed poor levels of performance with ketones compared with their performance with aldehydes and keto acids. As an example, the conversion of acetone was measured to be less than 0.04% relative to that for 2-oxobutyrate and propionaldehyde (8), which were the preferred keto acid and aldehyde substrates, respectively. A thermodynamic limitation for the amination of ketones is a common characteristic of ω-TAs (9). For instance, the k_{cat}/K_m ratio of the ω-TA from *Ochrobactrum arthropi* for acetophenone was only 0.0004% relative to that for pyruvate (10). Also, ω-TAs from *Parococcus denitrificans* and *Chromobacterium violaceum* showed from 0.015% to 0.083% relative activities for ketones compared to the values for α-keto acids and aldehydes (9). This prompted the research to create, by active-site engineering, ω-TA variants displaying improved capacity for the synthesis of chiral amines from bulky ketones. By applying this procedure, a 105-fold activity improvement for the conversion of butyrophenone was achieved for the ω-TA from *O. arthropi*, although the relative activity compared to the value for aldehyde was still low (from 1.4 to 11.3% relative activity) (10).

Finding new ω-TAs displaying high capability for the conversion of ketones, in combination with good stability and preferably stringent (*R*) or (*S*) selectivity (11–13), is thus a priority for the synthesization of pharmaceutically valuable chiral amines (9). However, their discovery is limited, most likely due to a lack of suitable screening methods at large scale. Recently, new assays for high-throughput screening of ATAs in liquid or solid phase were described (14, 15). By adapting these methods to screen a large collection of clone libraries generated from environmental DNA of diverse origins, we successfully identified 10 genes encoding presumptive ATAs of the class III ω-TA family. These genes were expressed in fusion proteins fused to polyhistidine (His) affinity tags, purified by immobilized metal affinity chromatography, and characterized. The results presented here illustrate the benefits of the methods herein applied to screen for ω-TAs using metagenomics. The extensive analysis of their substrate spectra allowed the identification of those capable of converting bulky ketones and bulky amines, as well as environmentally relevant amines like putrescine. Finally, the application of sequence and 3-dimensional-model analyses shed new lights on the molecular determinants of their substrate specificities and stereochemistry. The present study may help future bioprospecting and engineering programs to identify and design class III ω-TA family proteins converting bulky ketones and bulky amines with stringent (*R*) or (*S*) stereospecificity.

RESULTS

Gene selection by naive screens. Recently, a large set of metagenomic fosmid libraries from microbial communities inhabiting 28 geographically distinct environmental sites has been created and subjected to naive screen for esterase activity (16). In accordance with biological diversity and activity success rates, libraries from microbial communities inhabiting 10 of these sites (for details, see Materials and Methods) were chosen as starting points for screening new ω-TA-encoding sequences. We applied to all libraries two distinct naive agar-based screen methods, utilizing the two amine donors 2-(4-nitrophenyl)ethan-1-amine and *o*-xylylenediamine hydrochloride (see Fig. S1 in the supplemental material), which after transfer reactions render a colored product that can be identified by visual inspection (14, 15). About half a million pCCFOS1 fosmid clones (nearly 18 Gbp) from libraries generated from environmental DNA and 4,400 plasmid-based (pCR-XL-TOPO) clones from a *Pseudomonas oleovorans* genomic DNA library were screened for ATA activity using both agar-based screens. We identified a total of 10 positive clones active against 2-(4-nitrophenyl)ethan-1-amine, 3 of which were also active against *o*-xylylenediamine hydrochloride. They were recovered from clone libraries created from two chronically polluted marine sediment samples, an acidic beach pool, and the *P. oleovorans* genome (see the legend to Table S1) (16–18). This means a ratio of circa 1 positive result per 50,000 clones tested; note that this ratio is an indicator of the abundance of enzyme activities in metagenomes (5, 19). The 10 positive fosmid/plasmid inserts were fully sequenced using the Illumina MiSeq sequencing system. From the sequence data, 10 candidate genes encoding putative ω-TAs, one per insert, were identified.

Analysis of candidates at the protein sequence level. As shown by the results in Table S1 in the supplemental material, the deduced molecular mass and estimated isoelectric point (pI) values for the amino acid sequences comprising the 10 ω-TAs (designated TR_1 to TR_{10} based on the code TR, which refers to transaminase) ranged from 48.4 to 50.8 kDa and from 5.4 to 6.1, respectively. A comprehensive analysis of the TBLASTX results (20) indicated that putative proteins exhibited amino acid sequence identities ranging from 84% to 100% to sequences of uncharacterized homologous proteins in nonredundant public databases and from 35.7% to 60.7% to sequences in Protein Data Bank (PDB). As determined by Matcher (EMBOSS package), the pairwise amino acid sequence identities between the 10 selected sequences ranged from circa 33% to 99% (see Table S2A). TR_3 and TR_7 (98.9%), TR_4 and TR_5 (94.5%), and TR_9 and TR_{10} (92.9%) shared the highest sequence similarities. Sequence analysis and TBLASTX (20) categorized all sequences within the class III ω-TA family (21). It is interesting to note that the naive assays employed led to the identification of only class III ω-TAs and none of the other 5 groups, referred to as classes I, II, IV, V and VI (1). It may be due to the fact that compared to ATAs from other groups, ω-TAs transfer an amino group from an amine donor onto a carbonyl moiety of an amine acceptor, in which process at least one of the two substances is not an α-amino acid or an α-keto acid (1). Any of these substrates was used in the naive screens, which may explain why the screening with the two amine donors 2-(4-nitrophenyl)ethan-1-amine and *o*-xylylenediamine hydrochloride produced a bias toward class III ω-TAs.

Taken together, these data suggest that there is a large divergence at the sequence level within the identified class III ω-TA sequences and that the diversity is not dominated by a particular type of protein or highly similar clusters of proteins but consists of diverse nonredundant ω-TA sequences. Phylogenetic binning (22, 23) of the sequences encoding presumptive class III ω-TAs further revealed that they most likely originated from bacteria of the *Rhodobacteraceae* family (5 sequences) and the *Pseudomonas* (3 sequences), *Acidihalobacter* (1 sequence), and *Amphritea* (1 sequence) genera (see Supporting Results and Fig. S2 in the supplemental material).

Recombinant protein production. The 10 candidates were cloned into pBXCH or pRhokHi-2 (with C-terminal His tags) or into pBHXN3 (N-terminal His tags) (Fig. S1); these expression vectors allow high-expression yields for genes of environmental and

microbial origins (16, 24, 25). *Escherichia coli* strains MC1061 (when using pBXCH and pBHXN3 vectors) and DH5α (for pRhokHi-2) were used as expression hosts. The preferred construct design for each enzyme was selected on the basis of the protein yield obtained, which is out of the scope of the present study and, thus, not shown. The average yields of the ω-TAs in *E. coli* heterologous expression systems were approximately 32 mg purified protein per liter of culture. The 10 hexahistidine-tagged candidates were purified from lysed cells by immobilized metal affinity chromatography using Ni-nitrilotriacetic acid (NTA) technology (Fig. S3). The substrate profiles and properties of purified ω-TAs were further analyzed.

Acceptor profiling: preference for bulky ketones. The acceptor profiling was performed by using the colorimetric liquid assay reported by Baud et al. (14) with minor modifications (see Materials and Methods). We used one primary amine donor [2-(4-nitrophenyl)ethan-1-amine] and a set of 18 chemically and structurally distinct aldehydes, ketones, and keto acids as amine acceptors (see Scheme S1 in the supplemental material). Alkyl ketones of different lengths and large aromatic ketones and aldehydes were also included, since a larger group adjacent to the aldehyde or ketone group increases the difficulty of amine transfer. Reaction mixtures obtained in all cases were analyzed by electrospray ionization mass spectrometry (ESI-MS) (see Materials and Methods), and the existence of reaction products was confirmed in each case (see Supporting Results in the supplemental material).

The relative (%) specific activities obtained with the best substrates for all 10 ω-TAs for each of the 18 acceptors are summarized in Fig. 1; for specific activity (U g^{-1} protein) raw data, see Table S2B. We first observed that according to the data for the best acceptor, TR_1, with a maximum specific activity of 95.7 U g^{-1}, was also the least active enzyme, with the other ω-TAs having values ranging from 972 U g^{-1} to 179 U g^{-1} (Table S2B). The specific activities for the best substrates are in the range previously reported for ω-TAs (12, 26).

Regarding acceptor range, only 6 of 18 amine acceptors were converted by all 10 ω-TAs (Fig. 1). We found that TR_1 was characterized by a restricted acceptor spectrum, only capable of using as acceptors the keto acids glyoxylic acid, levulinic acid, and α-ketoglutaric acid (nonpreferred acceptors) and, to a much greater extent, the aldehydes hexanal (preferred acceptor) and benzaldehyde and the ketone 2-hexanone. No capacity of TR_1 to use any of the other linear and bulky ketones tested was detected under the assay conditions. All of the other 9 ω-TAs showed broader acceptor spectra and were characterized by significantly higher preferences for ketones than for aldehydes and keto acids. Thus, on the basis of specific activity determinations (U g^{-1}), the relative activities for amine transfer to the best accepted ketone compared to the preferred aldehydes or keto acids were found to be 50.4% (for TR_1), 96.1% (for TR_2), 9.0% (for TR_3), 90.3% (for TR_4), 112.5% (for TR_5), 1.1% (for TR_6), 84.3% (for TR_7), 14.9% (for TR_8), 173.4% (for TR_9), and 142.8% (for TR_{10}), as measured at 40°C and pH 7.5 (Table S2B). These percentages are significantly higher than those observed for previously reported ω-TAs (below 0.083%) (8–10).

Preferred acceptors for most ω-TAs were hexanal, benzaldehyde, or the bulky ketone 1-*N*-Boc-3-pyrrolidinone, depending on the transaminase (Fig. 1). Other bulky ketones, particularly 2-acetylpyridine and 2-acetylpyrazine, were well accepted by all ω-TAs, albeit they were not the preferred acceptors. Concerning the use of aliphatic ketones as acceptors, we found that increases in the side chain length resulted in lower levels of conversion (2-hexanone $>$ 2-heptanone $>$ 2-nonanone). As shown by the results in Fig. 1, bulky ketones were converted at levels similar to or higher than aliphatic ketones and aldehydes. Concerning the use of ketoacids, TR_6 and TR_8 had a noticeably higher preference for glyoxylic acid than did the other ω-TAs.

The similar acceptor profiles for TR_9 and TR_{10} (Fig. 1) agreed with their 93% identity at the amino acid sequence level (Table S2A). However, TR_3 and TR_7, which shared 99% identity, differed in the capacity to use multiple acceptors, which was particularly noticeable for their capacity to use 1-*N*-Boc-3-pyrrolidinone (relative activities of 1.6%

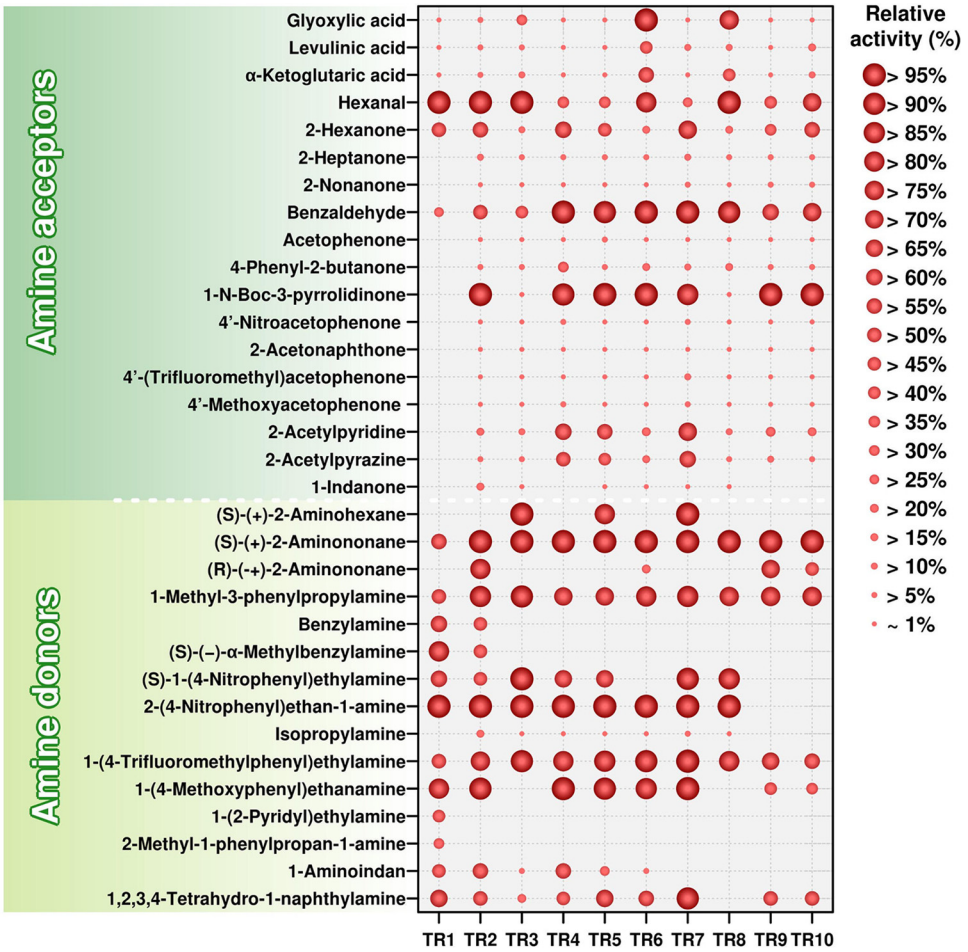


FIG 1 Substrate ranges of class III ω-TAs. Amine acceptors (top) and donors (bottom) tested are shown on the left. The identification code for each enzyme is shown on the bottom. This figure was created from data presented in Tables S2B and C in the supplemental material. 2-(4-Nitrophenyl)ethan-1-amine was used as the amine donor with the aldehydes, ketones, and keto acids listed as acceptors. Benzaldehyde was used as the acceptor with the amines listed as donors. (*R*)-(+)-Aminohexane, (*R*)-(+)-α-methylbenzylamine, and (*R*)-1-(4-nitrophenyl)ethylamine are not indicated because their conversion was below the detection limit under our assay conditions. The specific activity (U g^{-1}) at 40°C and pH 7.5 was determined as described in Materials and Methods, and the relative activity (%) obtained with the best acceptor (top) or donor (bottom) is indicated. The figure was created with the R language console. The structures of the acceptors and donors can be seen in Schemes S1 and S2 in the supplemental material.

for TR_3 and 84.31% for TR_7) and glyoxylic acid (relative activities of 28.8% for TR_3 and 1.04% for TR_7). TR_4 and TR_5, which were 94.5% identical, only showed an appreciable difference for the conversion of 4-phenyl-2-butanone (relative activities of 31.7% for TR_4 and 3.3% for TR_5); 1-indanone was only converted by TR_5, albeit at a significantly low relative activity (0.93%) compared to the value for its best acceptor. It was particularly noticeable that, excluding TR_1, which was unable to use 1-*N*-Boc-3-pyrrolidinone, TR_3 and TR_8 showed significantly low relative specific activities for this acceptor (1.6% for TR_3 and 0.3% for TR_8) compared to all of the other ω-TAs, for which this was one of the preferred acceptors (from 84.3% to 100% relative activity).

Donor profiling: preference for bulky amines. The ω-TAs were examined using benzaldehyde as a ketone acceptor and a set of 14 chemically and structurally distinct amine donors (see Scheme S2), including 4 pairs of enantiomers and the inexpensive isopropyl amine (Fig. 1; for raw data, see Table S2C). Instead of using the common gas chromatography (GC) method for detecting transamination products (12), we adapted the colorimetric assay described by Baud et al. (14). Briefly, we detected the amount of benzaldehyde that remained after the reaction compared to the amounts remaining

after control reactions in the absence of the amine donor and in the absence of enzymes (see Materials and Methods, as well as Scheme S2). The reaction mixtures obtained in all cases were also analyzed by ESI-MS (see Materials and Methods), and the formation of reaction products confirmed (see Supporting Results in the supplemental material).

We first observed that, using benzaldehyde and the preferred amine donor, the maximum activity ranged from 22.8 to 841.5 U g^{-1} (Table S2C), which is in the range found for other reported ω-TAs (12, 27). Regarding specificity, mayor differences in the capacity to use the amine donors were noticed, although all of them were characterized by an ample amine spectrum (Fig. 1). Only 4 of 14 amine donors were converted by all 10 ω-TAs, including (*S*)-(+)-2-aminononane and the bulky amines 1-methyl-3-phenylpropylamine, 1-(4-trifluoromethylphenyl)ethylamine, and 1-(4-methoxyphenyl)ethanamine, which were among the preferred amines in most cases. The other 10 amines were distinctly converted, highlighting the following major differences in the capacity to convert bulky amines. TR_1 was the only one using 1-(2-pyridyl)ethylamine and 2-methyl-1-phenylpropan-1-amine as amine donors; the capacity to use these bulky amines contrasts with the low capacity of TR_1 to use bulky ketones as acceptors (Fig. 1). To date, only a few (*S*)-selective ω-TAs, but not (*R*)-selective ones, have been identified as being capable of accepting a propyl group of amine substrates in the so-called S pocket (28). The fact that TR_1, which converts (*R*) and (*S*) amines (see below), was capable of using as an amine donor 2-methyl-1-phenylpropan-1-amine, which contains an isopropyl group, is thus an unusual feature for this enzyme. Benzylamine and (*S*)-(−)-α-methylbenzylamine were the only donor substrates for TR_1 and TR_2, (*S*)-1-(4-nitrophenyl)ethylamine was not accepted by TR_9 and TR_{10}, and 1,2,3,4-tetrahydro-1-naphthylamine was not converted by TR_8. Concerning the use of aliphatic amines, we found that larger amines were preferred. Indeed, all of the ω-TAs converted (*S*)-(+)-2-aminononane, but only 3 (TR_3, TR_5, and TR_7) of 10 converted the shorter (*S*)-(+)-2-aminohexane. It was particularly noticeable that all but TR_1, TR_9, and TR_{10} were capable of using isopropyl amine as a donor substrate, although it was not the preferred donor (Fig. 1).

In agreement with their high sequence identity (Table S2A), TR_9 and TR_{10} showed similar donor profiles. TR_3 and TR_7 (99% identical) were capable of using 10 amine donors, with a notable difference in their capacity to convert 1-(4-methoxyphenyl)-ethanamine, which was not converted by TR_3 but was one of the preferred donors for TR_7 (relative activity of 96%). Finally, the 94.5% identical TR_4 and TR_5 converted 10 amines, with major differences only noticed in the conversion of (*S*)-(+)-2-aminohexane, which was not converted by TR_4 but was one of the preferred donors for TR_5 (relative activity of 80.3%).

Sequence similarity and GNN analysis. Sequence similarity network (SSN) analysis of the TR_1 to TR_{10} sequences was performed with the aminotransferase class III collection of the InterPro database (IPR005814; 55,400 entries). The network is shown in Fig. 2A and B, where each node represents entries with 60% or higher sequence similarity and the lengths of the edges correlate with the dissimilarity of the connected sequences represented by the nodes (organic layout). Although the pairwise amino acid sequence identities between the 10 selected sequences ranged from circa 33% to 99%, they were all located in a central color cluster. A secondary SSN built from the first 500 BLAST hits against each sequence query (Fig. S4A) was used for building a genome neighborhood network (GNN). Using a window of 5 open reading frames (ORFs) upstream and downstream from the candidate TR_1 to TR_{10} genes, the genomic context of each of the target sequences was inferred.

The GNN for the TR_2, TR_9, and TR_{10} cluster is shown in Fig. S4B. The GNN was highly homogenous with regard to the ORF architecture (Fig. 2C), consisting of a putrescine-binding periplasmic protein (PotF), a spermidine/putrescine import ATP-binding protein (PotA), a putrescine transport permease, a Gln-synt_C (glutamine synthetase

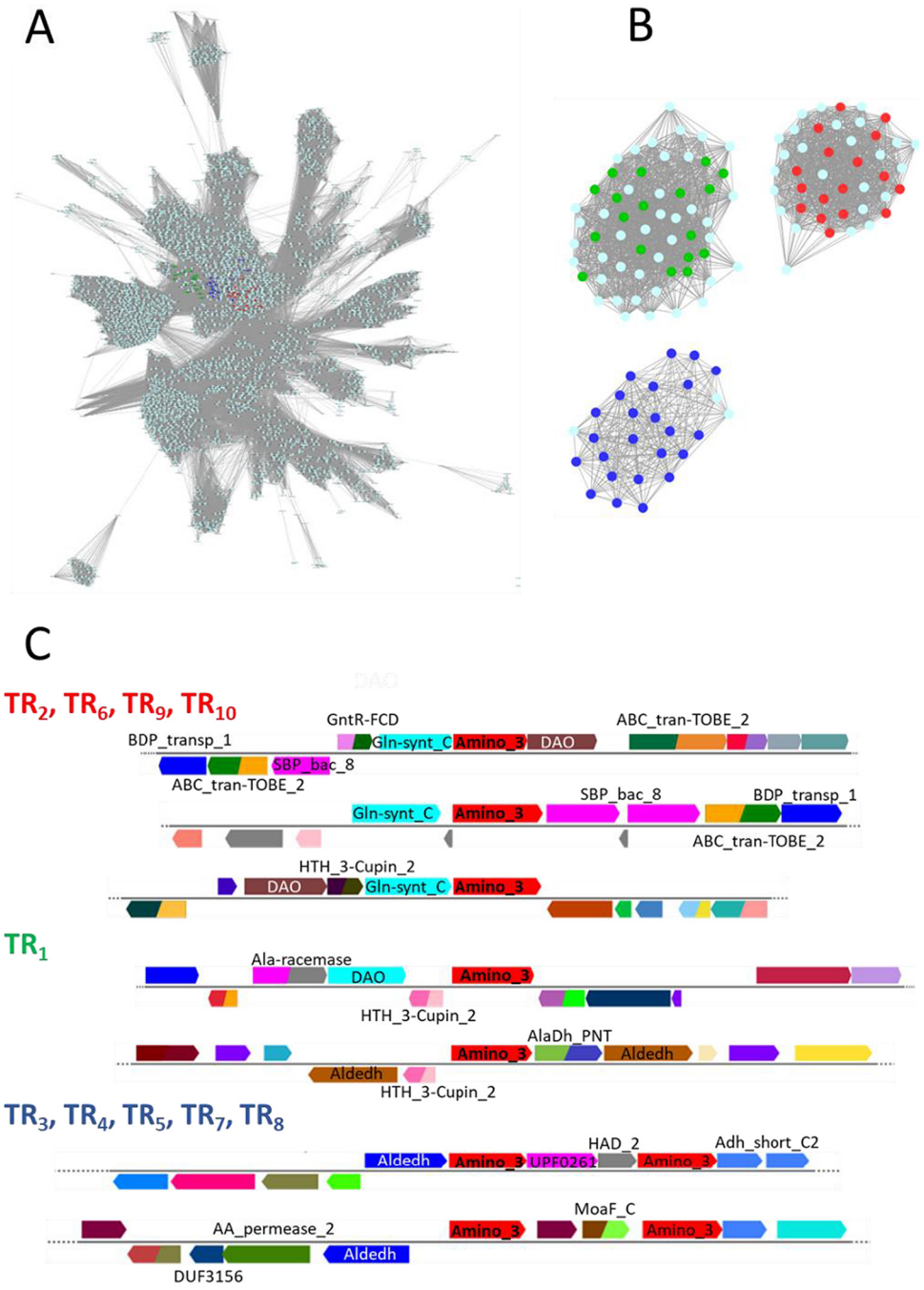


FIG 2 Genomic context analysis. (A) Sequence similarity network (SSN) analysis of transaminase family III (InterPro entry IPR005814) generated using the UniRef90 database with an E value of $10e^{-70}$. The nodes containing the 500 BLAST hits to the different TR candidates are in color (red for TR_2, TR_6, TR_9, and TR_{10}, green for TR_1, and blue for TR_3, TR_4, TR_5, TR_7, and TR_8). These three clusters segregate completely with an E value of $10e^{-100}$. (B) Enlargements of ω-TA clusters with an E value of $10e^{-120}$. Genome neighborhood networks (GNNs) were built using a window of 5 ORFs upstream and downstream from the candidate TR gene and a 20% threshold for cooccurrence. (C) The most common gene architectures for the three clusters are shown with Pfam annotations. Note that the color code may vary for ORFs with the same annotation.

catalytic domain) that is likely to act as a gamma-glutamyl-putrescine synthetase, which ligates glutamate and putrescine, and an aminotransferase (Fig. 2C). Together, TR_2, TR_9, and TR_{10} might be involved in the catabolism of polyamines, more specifically putrescine (Fig. 3). The GNN for the TR_6 cluster (Fig. S4B) and the ORF architecture (Fig. 2C) revealed that, despite the pattern variations, the same elements are present: the

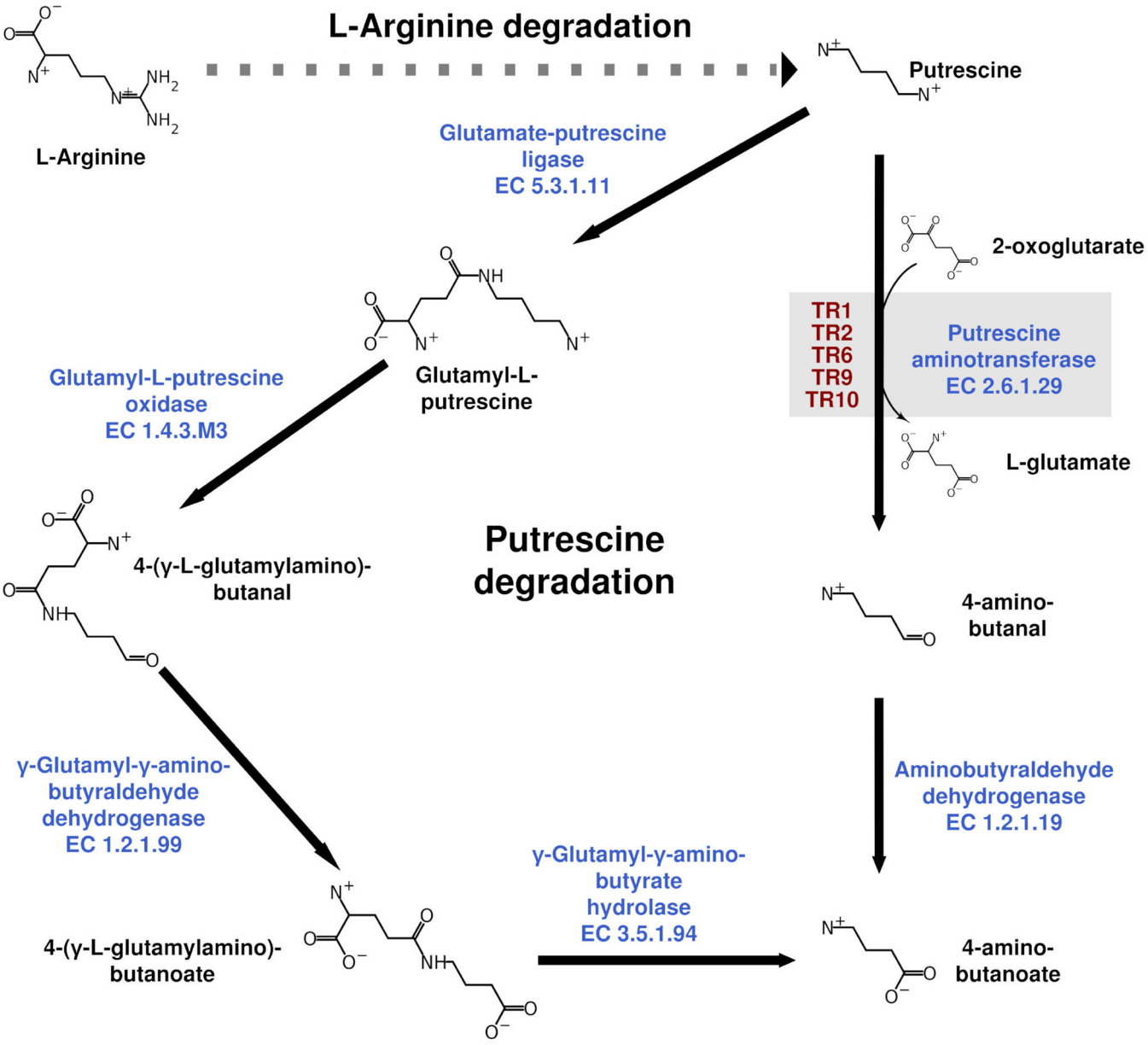


FIG 3 Reconstruction of the putrescine catabolic pathways in which the TR_1, TR_2, TR_6, TR_9, and TR_{10} class III ω-TAs may be implicated.

transporter proteins, Gln-synt_C, the aminotransferase, and an oxidase mostly related to glycine/D-amino acid oxidase or gamma-glutamyl-putrescine oxidoreductase. In summary, TR_2, TR_6, TR_9, and TR_{10} seem to be related to putrescine catabolism (Fig. S4D). In this pathway, the only transaminase occurs after obtaining 4-aminobutanoate (also known as GABA), which serves as the substrate for the transaminase (Fig. S4C). However, it is plausible that putrescine can also be a direct substrate for the transaminase; this is an alternative catabolic pathway that generates glutamate and 4-aminobutanal that is further oxidized to GABA (Fig. 3).

The GNN clusters (Fig. S4B) and genomic contexts (Fig. 2C) for the other transaminases (TR_1, TR_3 to TR_5, TR_7, and TR_8) do show many common elements, but there are also significant variations. The major difference is that genes encoding both the putrescine-ABC transporter system and the glutamate:putrescine ligase are missing, but others are present in the vicinity, including genes for an aldehyde dehydrogenase family protein (Aldedh), an aminotransferase 3, and another type of transporter

TABLE 1 Capacities of the class III ω-TAs to use putrescine as an amine donor

ω-TA	Sp act (U g^{-1})[a]
TR_1	56.8 ± 5.4
TR_2	1,190 ± 14.0
TR_3	ND
TR_4	ND
TR_5	ND
TR_6	543.6 ± 4.1
TR_7	ND
TR_8	ND
TR_9	461.7 ± 11.0
TR_{10}	476.7 ± . 9.8

[a]Benzaldehyde (10 mM) was used as the acceptor and putrescine (20 mM) as the donor. Specific activities, expressed as units per gram (U g^{-1}) at 40°C and pH 7.5, were determined as described in Materials and Methods. ND, not determined.

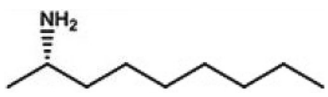
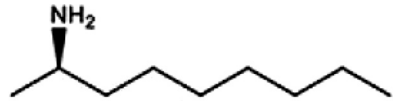
(amino acid permease). The SSN analysis suggests that most entries corresponding to Aldedh are annotated as 5-carboxymethyl-2-hydroxymuconate dehydrogenase (tyrosine metabolism), betaine-aldehyde dehydrogenase (glycine, serine, and threonine metabolism), succinate-semialdehyde dehydrogenase (putrescine and GABA degradation), gamma-glutamyl-gamma-aminobutyraldehyde dehydrogenase (putrescine catabolism), methylmalonate-semialdehyde dehydrogenase (inositol metabolism, valine, leucine, and isoleucine degradation and propionate metabolism), or phenylacetaldehyde dehydrogenase (phenylalanine metabolism). Together, we may suggest that these transaminase genes are related to amino acid catabolism but not putrescine catabolism, because of the absence of an ABC transporter and Gln-synt_C. However, the possibility that these genes are implicated in the catabolism of putrescine cannot be ruled out: the transport is through a permease different from the ABC type, and the two aminotransferases and the Aldedh are related to the catabolism of putrescine via 4-aminobutanal and GABA.

Based on the SSN and GNN analyses, the roles of TR_1 to TR_{10} *in vivo* in the catabolism of putrescine may be investigated. Potential substrates may be putrescine and GABA. They were tested as amine donors using benzaldehyde as the acceptor. The progress of the reaction was followed by measuring the amount of benzaldehyde remaining at the end of the reaction (see Materials and Methods and Scheme S2) and by ESI-MS (see Materials and Methods), respectively.

The data provided in Table 1 confirmed that TR_1, TR_2, TR_6, TR_9, and TR_{10} are capable of using putrescine as an amine donor when using benzaldehyde as the acceptor. None of the enzymes converted GABA. This suggests their implication *in vivo* in putrescine catabolism via the formation of 4-aminobutanal (Fig. 3). In contrast, TR_3 to TR_5, TR_7, and TR_8 did not use putrescine or GABA as an amine donor (Table 1), so it is most likely that they may participate in the catabolism of polyamines other than putrescine. Overall, the results agree with the SSN and GNN results.

Evaluation of enantioselectivity. The selectivity toward chiral amines was first determined by means of the selectivity factor, calculated as the ratio of specific activities (U g^{-1}) of the preferred over the nonpreferred chiral amines when both (*R*) and (*S*) amines were tested separately (see Scheme S3). It should be mentioned that these apparent values may not correspond to true selectivity or enantiomeric factors calculated when the enzyme is confronted with a racemic mixture, because the rates of transamination of the enantiomers were measured separately; nevertheless, recent studies for other classes of enzymes have clearly demonstrated that apparent and true selectivity values closely match each other (29). Briefly, the specific activities for (*S*)-(+)-2-aminohexane, (*R*)-(−)-2-aminohexane, (*S*)-(+)-2-aminononane, (*R*)-(−)-2-aminononane, (*S*)-(−)-α-methylbenzylamine, (*R*)-(+)-α-methylbenzylamine, (*S*)-1-(4-nitrophenyl)ethylamine, and (*R*)-1-(4-nitrophenyl)ethylamine were determined by applying a colorimetric assay in which benzaldehyde was used as the amine acceptor, as described above.

TABLE 2 Enantioselectivities of selected class III ω-TAs

ω-TA	Sp act (U g^{-1}) by colorimetric tests[a]		Selectivity according to asymmetric synthesis	
	(*S*)-(+)-2-Aminononane	(*R*)-(−)-2-Aminononane	Conversion (%)[b]	%eeP[b]
TR_1	12.7 ± 0.4	0	10.7 ± 1.1	100 (*S*)
TR_2	310.8 ± 0.70	255.3 ± 5.6	90.7 ± 9.5	49 ± 5.1 (*R*)
TR_3	386.1 ± 1.3	0	98.3 ± 3.4	100 (*S*)
TR_4	217.6 ± 0.2	0	47.3 ± 2.1	100 (*S*)
TR_5	524.1 ± 7.4	0	71.4 ± 3.0	100 (*S*)
TR_6	527.9 ± 0.6	98.6 ± 1.8	83.3 ± 2.3	35.6 ± 2.5 (*R*)
TR_7	814.5 ± 3.1	0	99.2 ± 4.8	100 (*S*)
TR_8	673.8 ± 3.5	0	98.9 ± 1.4	100 (*S*)
TR_9	149.3 ± 2.8	10.2 ± 0.6	60.1 ± 7.0	47.2 ± 6.0 (*R*)
TR_{10}	119.5 ± 4.5	5.34 ± 0.03	62.2 ± 4.0	48.6 ± 2.2 (*R*)
Chemical structure				

[a]Specific activities (U g^{-1}) measured in triplicate (with standard deviations shown) at 40°C and pH 7.5 after 60-min reactions, using benzaldehyde as the amine donor, as described in Materials and Methods.

[b]Percent conversion and percent enantiomeric excess of product (%eeP) according to the asymmetric synthesis of (*R*)-aminononane and (*S*)-aminononane. In the case of ω-TAs producing (*R*) and (*S*) amines, the %eeP for the (*R*) amine is shown; for those only producing (*S*) amine, the %eeP for the (*S*) amine is shown. The formation of the reaction product 2-nonane ($C_9H_{18}O$) was confirmed by ESI-MS (the theoretical exact mass was determined to be 142.1358 Da, and the experimental value obtained by ESI-MS was 142.1370).

As shown by the results in Fig. 1 and Table 2, we found that under our assay conditions, 6 ω-TAs were stringently (*S*) selective, with no appreciable activity for any of the (*R*) amines tested. TR_2, TR_6, TR_9, and TR_{10} were capable of converting both (*R*) and (*S*) enantiomers of 2-aminononane, with the (*S*) enantiomer being preferred (Fig. 1). For these 4 ω-TAs, the selectivity factors, calculated as the ratio of specific activities (U g^{-1}) of the preferred [(*S*)-(+)-2-aminononane] over the nonpreferred [(*R*)-(−)-2-aminononane] chiral amine, ranged from ~1.2 to 5.4 (Table 2). The selectivity, as calculated by kinetic resolution and asymmetric synthesis and gas chromatography (see Scheme S3 and Fig. S5), also confirm the results of the colorimetric assays (Table 2).

It has previously been demonstrated that cosolvents (i.e., 30% dimethyl sulfoxide [DMSO]) can change the selectivity of ω-TAs (13). For this reason, the apparent selectivity factors for TR_2, TR_6, TR_9, and TR_{10} were evaluated by colorimetric assays in the presence of 30% DMSO, a concentration at which the enzymes retain more than 67% of their activity (see below). No statistically significant changes in apparent selectivity factors were observed against (*S*)- and (*R*)-2-aminononane (data not shown).

Optimal parameters for activity. Using benzaldehyde and 2-(4-nitrophenyl)ethan-1-amine as the acceptor and donor, respectively, the purified ω-TAs were found to be most active at temperatures ranging from 45 to 65°C (Fig. 4). TR_3, TR_6, TR_7, and TR_8 were the most active at the high temperature range (60 to 65°C). A tolerance for organic solvents increases the industrial utility of enzymes. The amination of benzaldehyde with 2-(4-nitrophenyl)ethan-1-amine was tested at different concentrations (from 5% to 50% [vol/vol]) of water-miscible solvents, namely, methanol, acetonitrile, DMSO, dimethyl acetamide (DMA), isopropanol, and acetone (Fig. 5). TR_2, TR_4, TR_5, and TR_{10} were most active in the presence of acetone (from 45.0 to 51.7% relative activity at 50% acetone), and TR_3, TR_6, TR_7, and TR_9 were most active in the presence of DMSO (circa 100%, 42%, 63%, and 44% relative activity at 50% DMSO). TR_8 retained more than 59% relative activity in the presence of DMA, DMSO, and isopropanol at 50% each.

Molecular determinants of substrate specificity. The extensive analysis of the substrate spectra showed a number of differences (Fig. 1), possibly a reflection of the divergence in steric constraints and active-site architecture, which may help to better understand the specificity and stereochemistry of the ω-TAs herein investigated.

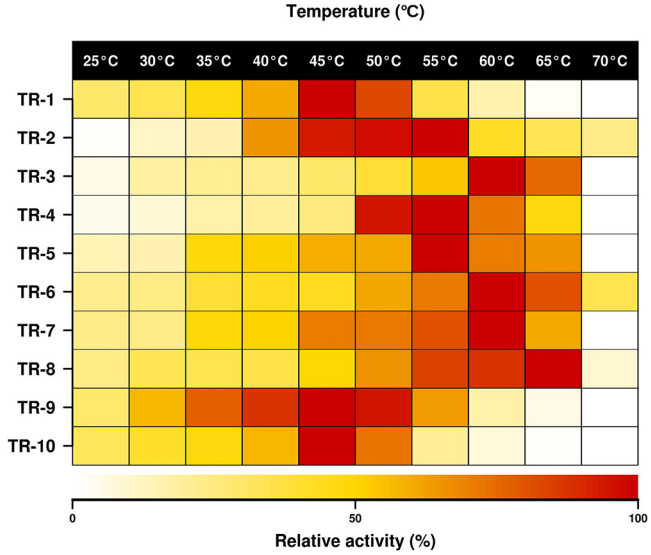


FIG 4 Temperature profiles for purified class III ω-TAs. The data represent the relative percentages of specific activity at pH 7.5, expressed as U g^{-1}, compared with the maximum activity when benzaldehyde and 2-(4-nitrophenyl)ethan-1-amine were used as the acceptor and donor, respectively. Full data sets are given in Table S2D.

An extensive analysis of the key active-site residues determining the substrate specificity in previously reported ω-TAs (30) through multiple sequence alignments and three-dimensional structural models (Fig. S6 to S10) was performed. This analysis was done in relation to substrate specificity data (Fig. 1). We first found that residues previously suggested to play a role in substrate recognition (Fig. S7) (30) may not have a role in determining the substrate spectra and preference for bulky substrates in the 10 ω-TAs herein described (see Supporting Results in the supplemental material). Rather, most likely, a major molecular determinant of the capacity of class III ω-TAs herein described to use bulky ketones as acceptors is the presence of a hairpin region proximal to the highly conserved Arg414 (following the numeration in the ω-TA from *Pseudomonas putida* [PDB 3A8U]) (Fig. 6), an extensively characterized ω-TA whose X-ray structure has been determined (30). This hairpin region is present in TR_2 to TR_{10} but not TR_1 nor the PDB 3A8U sequence (Fig. 7). We hypothesize that the presence of this hairpin, the residues comprising it (Fig. 7), and the nature of 3 amino acids in its proximity play a role in determining the orientation of the conserved Arg414 and the access to and positioning of bulkier ketones in relation to the L pocket to an extent greater than that previously thought for the conserved Arg414 (30). This can be seen by the results in Fig. 7 and, in more detail, in Fig. S8. In addition to this, as shown in Supporting Results and Fig. S7 and S10 in the supplemental material, most likely the positioning of the conserved Ser231 (following the numeration in the ω-TA with PDB 3A8U) may additionally play a role in the capacity of class III ω-TAs to use amines with longer alkyl substituents, to a greater extent than previously thought (27).

TR_2, TR_6, TR_9, and TR_{10} were the only class III ω-TAs herein reported that were capable of converting (*R*) amines, although their low selectivity cannot give access to highly optically pure amines. In order to increase the possibility of designing (*R*)-selective variants, an inspection of the literature was first undertaken. Note that (*R*)-selective amine transaminases have been only reported within the PLP-dependent fold class IV ω-TAs, which were found by applying *in silico* mining approaches (28, 29). Sequence alignment of the TR_1 to TR_{10} sequences with sequences of (*R*)-selective PLP-dependent fold class IV proteins (28) revealed that the conserved residues implicated in the recognition of (*R*) amines in class IV ω-TAs were found in the class III ω-TAs herein reported (Fig. S11). This suggests that residues other than those reported for

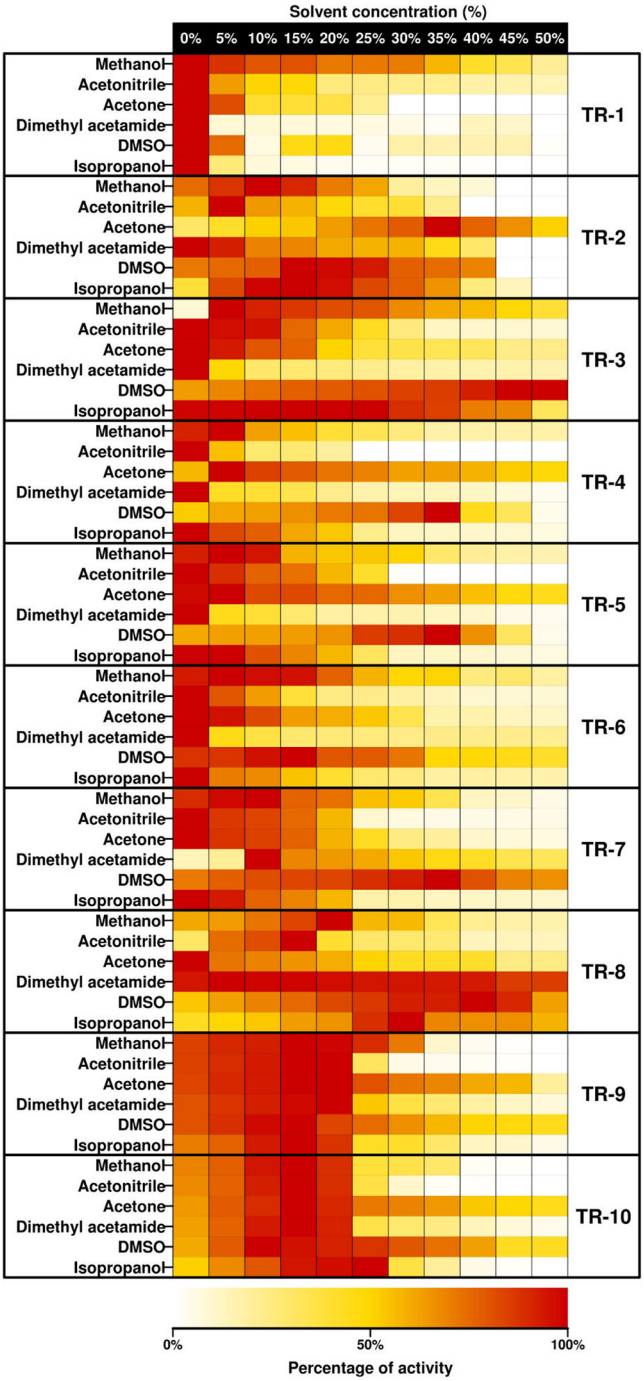


FIG 5 Solvent resistance profiles for purified class III ω-TAs. The data represent the relative percentages of specific activity at pH 7.5, expressed as U g^{-1}, compared with the maximum activity when benzaldehyde and 2-(4-nitrophenyl)ethan-1-amine were used as the acceptor and donor, respectively. Full data sets are given in Table S2E.

class IV ω-TAs are determining the capacity of class III ω-TAs to convert (*R*) amines. The analysis of sequence alignments and three-dimensional models revealed that TR_2, TR_6, TR_9, and TR_{10}, in contrast to the other class III ω-TAs herein reported, have in common the presence of a large pocket volume, the presence of a hairpin region close to the conserved Arg414, and an outward orientation of Arg414 (Fig. 7; Fig. S6 to S8). We suggest that, those structural elements aside, others yet to be determined may be responsible for the capacity of both enantiomers to be positioned in the active sites of

```
                 430        440        450        460        470        480
            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
3A8U        VRPFEAGMAL WKAG------ -FYVRFGGDT LQFGPTFNSK PQDLDRLFDA VGEVLNKLLD
TR1         ARGYEVFDSC FWHG------ -AMIRCTGDI LAMSPPLIAE KQDLDRLIEI VSKVIKQTA-
TR2         RFEEPG-KVG SLCRDLSVKN GLVMRAVGGT MIISPPLVLS REQVDELIDK ARRTLDETHK
TR3         TPMDNET--M KRIHQTAYEA GAMVRLGAHN VLMSPPLTIS EAEVNTILTA LDAGFSAA--
TR4         TAAAANT--V SMVFNTAYEA GAMVRIGGNN LLMSPPLIIT EAEVDVILSA LDTGLSAVST
TR5         AAAAPGT--V AKAFNTAYDA GAMVRIGGNN LLMSPPLIIT EGEVDVILSA LDAGLTAASS
TR6         TFKAEAGTVG YICRERCFAN NLVMRHVGDR MIISPPLTLT RDEIDLLIER AWKSLDEGMA
TR7         TPMDGAT--M ARIHQTAYEA GAMVRLGAHN VLMSPPLTIS EAEVNTILTA LDAGFSAA--
TR8         TPIDPMGGFA NQIAAVALRE GVIARPVGTK IIISPPLTIG TEEVDKMVAA LLQGFTEVDR
TR9         RFSGDV-GVG MVCRGHCFNN GLIMRAVGDT MIIAPPLVIS QTEVDELVEK ARKCLDLTWE
TR10        RYPGDK-AVG MICRGHCFNN GLIMRAVGDT MIIAPPLVIS QAEVDELVEK ARKCLDLTLR
A0A1H1SVF6  ARGYEVFDGC FWQG------ -AMIRCTGDI LAMSPPLIAE KQDLDQLIDI VSKVIKQTA-
A0A1H6ARZ3  ARGYEVFDGC FWEG------ -AMVRCSGDI IAMSPPLTVE KAELDRLVDT LATVIRRTA-
A0A1H2HXV0  TRGYDVFEEC FWQG------ -AMVRCTGDI IAMSPPLIAE KEHIDQLIEI LGRVINKAN-
A0A1H1YCL7  TRGYDVFESC YWEGD----- -VMVRCTGDI IAMSPPLITE KTHIDQLVEA LGKVIKRTQ-
A0A078MFR8  TRGYEVFEGC FWEG------ -LMVRCTGDT LALSPPLTID HAHIDRIMDT LGRVIRRTA-
L8MTF3      KRGFQVFEQC FHDG------ -LMVRVTGDT IAMSPPLIVE KEQIDTLVGT LADSIRKAA-
S6AXN9      KRGFQVFEQC FHDG------ -VMVRVTGDT IAMSPPLIVE KEQIDILVGK LADSIRKAA-
KX505388    SRGYDIMREA LKRG------ -LLIRLTGDT IALSPPLIIE PSHMDRIFHT LRDVLRTTA-
A0A1J5LK70  VFGSDAGTVG YICRERCFAN NLVMRHVGDR MIISPPLVIK PEEIDVLIER ARKSLDECYA
KX505389    KFPPEFKLGP KLEAATRRR- GVIVRCTPDG IIMAPPLTIT REECDVLIEA VAGALSDVLD
A0A0M9EFX0  TPMDMAT--M KKVHEATYQA GAMVRLGAHN VLMSPPLTIS EAEVNTILSA LDAGFSAA--
A0A081G376  EPIDPTNGYA NRVAAVAMRE GAIVRPVGTK IILSPTLTLT ETEVDKLTSA LKIAFQEVKA
```

FIG 6 Conservation of the highly conversed Arg414 residue (bold and underlined) examined by multiple sequence alignment of the class III ω-TAs herein reported, that of *Pseudomonas putida* (PDB 3A8U), and those included in Fig. S2. Only the region in the proximity of Arg414 is shown. The source of the numbering is A0A081G376. Conservation of other residues suggested to be implicated in substrate specificity and PLP binding is shown in Fig. S6 and S7. The residues constituting the hairpin region close to the conserved Arg414 (following numeration in PDB 3A8U) are underlined and in gray.

TR_2, TR_6, TR_9, and TR_{10} and, thus, determine the capacity of such enzymes to convert (*R*) amines.

DISCUSSION

In this study, we adapted two high-throughput screening methods to identify 10 class III ω-TA family proteins. The identified ω-TAs originated from bacteria from at least

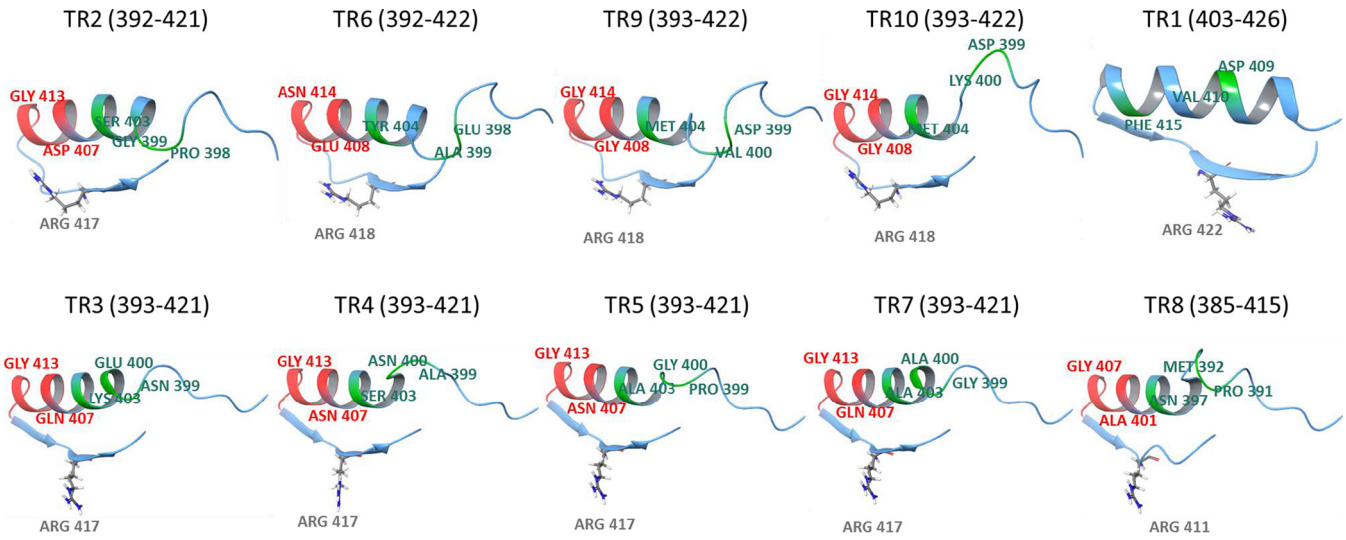


FIG 7 Structural investigation of the class III ω-TAs to elucidate the positioning of the hairpin region, located close to a highly conserved arginine (Arg414 in the ω-TA from *P. putida* [PDB 3A8U], as the reference); the orientation of the conserved arginine is also shown. Arg414 is located in a small hairpin (red) in the middle of a coil delimited by the two helices, one being shown. Only the region in the proximity of the conversed arginine is represented, with initial and end residues indicated in parentheses. The region present in TR_2 to TR_{10} (and also in PDB 3A8U [not shown in the figure]), but not in TR_1, is shown in red. Three additional residues located in the proximity of this region, which were found to play a role in determining the access to the L pocket (see Supporting Results), are shown in green. Although the modeling does not give an exact orientation of the Arg414 side chain, the presence of the hairpin and the residues located in its proximity seem to play a role in its orientation and also in the access to the active site, as can be seen in detail in Fig. S8. The PDB codes of templates used to create the models are shown in Fig. S7 and S8. (Top) The class III ω-TAs capable of degrading putrescine, with TR_2, TR_6, TR_9, and TR_{10} containing structural elements which differ from those of TR_1. (Bottom) The class III ω-TAs not capable of degrading putrescine. As shown, the orientation of the conserved arginine differs significantly among the groups of class III ω-TAs.

four different and divergent lineages: the *Pseudomonas*, *Acidihalobacter*, and *Amphritea* genera and the *Rhodobacteraceae* family. ω-TAs from the *Pseudomonas* genus (31) and *Rhodobacteraceae* family (32, 33) have been reported previously. However, this study examines ω-TAs derived from bacteria of the *Amphritea* (TR_8) and *Acidihalobacter* (TR_2) genera, which are bacterial groups rarely investigated from an enzymatic point of view (34, 35). Our results suggest that the class III ω-TA (TR_8) from the bacterium of the *Amphritea* genus is thermoactive (up to 65°C) and stable (it retained more than 35% activity in methanol, acetonitrile, and DMSO, each at a concentration of 50% [vol/vol]) and efficiently converts bulky substrates and (*S*) amines. The enzyme from the bacterium of the *Acidihalobacter* genus was capable of accepting bulky substrates as well as (*R*) and (*S*) amines. These features clearly suggest that those transaminases and their bacterial origins should be considered for chemical transformations in the future.

Particularly noticeable was the high level of performance with ketones in comparison to the results using aldehydes and keto acids, with specific activities for ketones being up to 173.4% relative to those of the best-accepted aldehyde and keto acid substrates. Such a preference for ketones is rarely observed for other native or engineered ω-TAs (8–10), which exemplifies the potential of bioprospecting programs to identify new enzymes for the amination of bulky ketones. Molecular determinants for this unusual specificity for bulky ketones were found and suggested, including a hairpin region proximal to the highly conserved Arg414 and residues in its proximity (Fig. 7) as a major determinant of the preference for bulky ketones and amines. Identifying transaminases containing this hairpin and applying rational or traditional protein engineering in the region may allow class III ω-TAs capable of converting bulky ketones and bulky amines to be designed. The conserved Ser231 was also found to be a major determinant of the preference for amines with longer alkyl substituents (Fig. S7).

The present study also reported 6 class III ω-TAs with a stringent (*S*) enantioselectivity. This is a common feature within most transaminases, such as the ω-TAs belonging to PLP type I fold, which are all specific toward the (*S*) enantiomer of their substrates. This also accounts for class III ω-TAs (30). Interestingly, we also reported four ATAs capable of acting toward (*S*) and (*R*) amines, with the (*S*) enantiomer being preferred. Note that both (*S*)- and (*R*)-selective ω-TAs have been found in distantly related families other than class III ω-TAs, namely those belonging to Pfam class IV transaminases with a PLP type IV fold (28, 36–40). Compared to the (*S*)-selective enzymes, the (*R*)-selective counterparts are less abundant and have been less studied. All (*R*)-specific class IV ω-TAs preferentially convert aliphatic substrates with high yields and high enantioselectivities [enantiomeric excess (ee) higher than 90% for (*R*) enantiomers], but the yields are significantly lower with aromatic substrates. Recently, the introduction of 27 mutations into a fold IV ATA allowed the substrate scope toward bulky substrates to be broadened (29). Therefore, four of the enzymes herein reported represent examples of the class III ω-TA family converting (*R*) amines, although they also convert (*S*) amines. Having a class III ω-TA acting toward (*S*) and (*R*) amines with low selectivity toward the latter may be more of a disadvantage than an advantage, as they cannot give access to highly optically pure amines. Although they are not (*R*) selective, the capacity to act efficiently toward (*R*) amines can be used as a starting point to apply rational design and protein engineering to design (*R*)-selective variants from the naturally occurring class III ω-TAs identified herein and, possibly, others. This study suggests a number of molecular determinants which may help in the rational design of such enzymes. These determinants include a large active-site pocket, the presence of a hairpin region close to the conserved Arg414, and the outward orientation of Arg414 (Fig. 7). Improving or even reversing the selectivity by single point mutations has also been shown by recent examples in other classes of ω-TA (11, 13), although engineering has been shown to lead to variants with low selectivity and it is not a universal effect for all substrates (11).

Taken together, this study reported examples of class III ω-TAs efficiently converting not only bulky ketones with stringent (*S*) stereochemistry but also one converting bulky ketones and bulky amines with a large alkyl substituent and a number converting bulky

ketones and both (*R*) and (*S*) amines. Four enzymes additionally retained significant activity up to 60 to 65°C, and five were stable in concentrations of up to 50% (vol/vol) of organic solvents. Altogether, the amine transaminases herein reported display biochemical properties that make them attractive candidates for a variety of chemical conversions and suggest future actions to design (*R*)-selective class III ω-TAs.

The characterization of the 10 class III ω-TAs herein described also allows their participation in polyamine catabolism, namely that of putrescine, a ubiquitous and important biological molecule (41), to be increased. We found that class III ω-TAs that contained the highly conserved Arg414 in an outward conformation (TR_2, TR_6, TR_9, and TR_{10}) (Fig. 7) were capable of degrading putrescine via the formation of 4-aminobutanal (Fig. 3). In the opposite case, class III ω-TAs in which Arg414 was in the inward conformation (TR_3 to TR_5, TR_7, and TR_8) (Fig. 7) were not able to degrade putrescine. TR_1 was also capable of degrading putrescine. Its sequence differs from those of TR_2 to TR_{10} by the absence of a hairpin region proximal to Arg414 (Fig. 7), causing an inward configuration of this residue, which is slightly differently oriented than those in TR_3 to TR_5, TR_7, and TR_8. Therefore, the orientation of the highly conserved Arg414 may be used not only as an indicator of the capacity of class III ω-TAs to degrade bulky ketones, bulky amines, and (*R*) and (*S*) amines but also as an indicator of the capacity to degrade putrescine. Note that in previously reported class III ω-TAs, the conserved Arg414 has also been shown to adopt different conformations, inward or outward (30). This orientation has been implicated in the recognition of carboxylate groups of keto acids and in determining the size of the large pocket (27). However, no environmental implication was suggested for the different orientations of the conserved Arg414. In this study, we found that the different orientations of the conserved Arg414 cannot, *per se*, be directly linked to the distinct capacity to convert keto acids over ketones, as was previously suggested (27, 30), since enzymes in which Arg414 is similarly oriented show marked differences in substrate preference and their capacity to use keto acids (Fig. 1). Rather, we found that the orientation can be linked to the capacity to degrade environmentally important biological polyamines (41), such as putrescine, as shown in this study (Fig. 3).

MATERIALS AND METHODS

General experimental procedures. All chemicals used for enzymatic tests were of the purest grade available and were purchased from Fluka-Aldrich-Sigma Chemical Co. (St. Louis, MO, USA). *E. coli* strains MC1061, a generous gift from Eric Geertsma, and DH5α were used for expressing TR_1 to TR_{10}.

Naive screens. The fosmid libraries used in the present study derived from 9 geographically distinct marine samples. They include samples from Port of Milazzo and Port of Messina (Sicily, Italy) (16, 18, 42), the Ancona harbor (Ancona, Italy), with uric acid and ammonium amendments (17), the Priolo Gargallo harbor (Syracuse, Italy) (18), the Arenzano harbor (Ligurian Sea, Genoa, Italy) (18), an acidic beach pool on Vulcano Island (Italy) (18, 42), the El Max site (Alexandria, Egypt) (18), Bizerte Lagoon (Tunisia) (18), and the Gulf of Aqaba (Red Sea, Jordan) (18). In all cases, DNA extraction and preparation of pCCFOS1 fosmid libraries were performed as described elsewhere (16, 18, 42). A genomic library of *P. oleovorans* strain DSM 1045 was constructed as described previously (43), with minor modifications. Briefly, genomic DNA of *P. oleovorans* DSM 1045 was isolated and fragmented by sonication, and appropriately sized fragments were then collected by gel extraction and end repaired (44). 5′-End phosphates were removed by using alkaline phosphatase (FastAP; Thermo Scientific) followed by DNA precipitation. 3′-End adenine overhangs were added by using *Taq* polymerase and cloned into the pCR-XL-TOPO vector according to the manufacturer's recommendations (TOPO XL PCR cloning kit; Invitrogen). The recombinant plasmids were then transformed into *E. coli* TOP10 cells by electroporation.

Clones were scored for their ability to perform transamination reactions by adapting a colorimetric assay (14). Briefly, the method is based on the use of 2-(4-nitrophenyl)ethan-1-amine as an amine donor that when converted into the corresponding aldehyde and subsequently deprotonated would give a highly conjugated structure with absorbance in the UV region and an orange/red precipitate. Fosmid clones were plated onto large (22.5- by 22.5-cm) petri plates with Luria-Bertani (LB) agar containing chloramphenicol (12.5 $\mu g\ ml^{-1}$) and induction solution (Epicentre Biotechnologies, Madison, WI, USA) in an amount recommended by the supplier to induce a high fosmid copy number. After overnight incubation at 37°C, the plates were overlaid with 40 ml of a solution of K_2HPO_4 buffer, pH 7.5 (100 mM), containing 0.4% (wt/vol) agar, to which the following chemicals were added immediately prior to use: PLP (10 mg, or 1.0 mM final concentration), the amine donor 2-(4-nitrophenyl)ethan-1-amine (202.64 mg, or 25 mM final concentration), and benzaldehyde (106.12 mg, or 25 mM final concentration) as the aldehyde acceptor. Note that the method can be adapted to any other ketone or aldehyde. Positive colonies in agar plates change to an orange/red color in 20 to 30 min when the colonies are overlaid with

the screening solution. Screens were also performed with *o*-xylylenediamine hydrochloride as the amine donor by adapting a colorimetric assay (15). Briefly, clones were plated on LB agar containing 12.5 μg ml^{-1} chloramphenicol (for screening the clone library from environmental sources) or 50 μg ml^{-1} kanamycin (for screening the *P. oleovorans* clone library). After overnight incubation at 37°C, clones were transferred to Whatman paper. A drop of reaction solution containing 5 mM *o*-xylylenediamine hydrochloride and 2.02 mM PLP in 100 mM K_2HPO_4 buffer, pH 8.0, was placed in the lid of a petri dish and covered with the colony-bearing Whatman paper. After the petri dish was sealed to prevent drying of the reaction solution, the plate was incubated at 30°C overnight. Positive clones with transaminase activity were identified by black coloration.

Positive clones containing presumptive transaminases were selected, and their DNA inserts were sequenced using a MiSeq sequencing system (Illumina, San Diego, CA, USA) with a 2 × 150-bp sequencing kit. Upon completion of sequencing, the reads were quality filtered and assembled to generate nonredundant metasequences, and genes were predicted and annotated as described previously (45). The sequences of the inserts of the plasmids containing TR_9 and TR_{10} genes were obtained from *P. oleovorans* genome data (46) after terminal sequencing of the plasmid insert (LGC Genomics GmbH, Berlin, Germany).

Gene expression. Two expression platforms were used. Codon-optimized synthetic versions of the TR_1 and TR_3 to TR_8 candidate genes were synthesized by GenScript (Hong Kong) and delivered in a customized pUC plasmid. These constructs were dissolved in sterile water upon arrival and used as delivery plasmid for subcloning by fragment exchange (FX) into expression vector pBXCH or pBXNH3 using *E. coli* MC1061 as a host (24, 47). Candidate genes for TR_2, TR_9, and TR_{10} were amplified from clonal DNA using gene-specific primers containing overhangs with restriction sites (NdeI/XhoI for TR_2 and NdeI/HindIII for TR_9 and TR_{10}) and cloned into expression vector pRhokHi-2 using *E. coli* MC1061 as the host (25).

Recombinant protein purification. All recombinant proteins were expressed with His tags and purified as follows. Briefly, selected *E. coli* clones that were found to convert the screening substrates were grown at 37°C on solid LB agar medium supplemented with the appropriate antibiotics (100 μg ml^{-1} ampicillin for pBXCH or pBXNH3 or 30 μg ml^{-1} kanamycin for pRhokHi-2). Single colonies were picked and used to inoculate 10 ml of LB broth supplemented with the appropriate antibiotic in a 0.25-liter flask. The cultures were then incubated at 37°C and 200 rpm overnight. Afterwards, 10 ml of this culture was used to inoculate 0.5 liter of LB medium plus antibiotic, which was then incubated to an optical density at 600 nm (OD_{600}) of approximately 0.7 (ranging from 0.55 to 0.75) at 37°C. The expression of TR_1 and TR_3 to TR_8 was induced by adding L-arabinose to a final concentration of 0.1%, followed by incubation for 16 h at 16°C. TR_2, TR_9, and TR_{10} were constitutively expressed using the same conditions (no inductor needed). In all cases, the cells were harvested by centrifugation at 5,000 × *g* for 15 min to yield a pellet of 2 to 3 g (wet weight). The wet cell pellet was frozen at −86°C overnight, thawed, and resuspended in 15 ml of 40 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.0. Lysonase bioprocessing reagent (Novagen, Darmstadt, Germany) was added (4 μl g^{-1} wet cells) and cells incubated for 60 min on ice with rotation. The cell suspension was sonicated for a total of 5 min and centrifuged at 15,000 × *g* for 15 min at 4°C, and the supernatant was retained. The soluble His-tagged proteins were purified at 4°C after binding to Ni-NTA His-Bind resin (Sigma Chemical Co., St. Louis, MO, USA), followed by extensive dialysis of the protein solutions against 100 mM K_2HPO_4 buffer, pH 7.5, by ultrafiltration through low-adsorption, hydrophilic, 10,000 (10K)-nominal-molecular-weight-cutoff membranes (regenerated cellulose, Amicon) and storage at 4°C. The proteins were further purified by gel filtration as described previously (48). Purity was assessed as >98% using SDS-PAGE analysis (49) in a Mini-PROTEAN electrophoresis system (Bio-Rad). The protein concentration was determined according to the Bradford method with bovine serum albumin as the standard (50).

Enzyme assays for determinations of acceptor substrates. Transaminase activity was assayed using 2-(4-nitrophenyl)ethan-1-amine and structurally diverse keto acids, aldehydes, and ketones in 96-well plates as previously described with some modifications (14). Note that K_2HPO_4 buffer was used, following recommendations described elsewhere (11, 14, 28). Briefly, assay reactions were conducted as follows. Prior to the assay, a solution of 25 mM amine donor 2-(4-nitrophenyl)ethan-1-amine and 1.0 mM cofactor PLP was first prepared in 100 mM K_2HPO_4 buffer, pH 7.5 (40 ml). A 400 mM acceptor (keto acid, aldehyde, or ketone) stock solution was prepared in acetonitrile or buffer, depending on solubility. Reaction assays, in 96-well microtiter plates, were started by adding 2.5 μl of a protein solution (stock solution, 10.0 mg/ml in 100 mM K_2HPO_4 buffer, pH 7.5) to an assay mixture containing 185 μl of PLP–2-(4-nitrophenyl)ethan-1-amine solution and 12.5 μl of acceptor stock solution. The final volume of the assay mixture was 200 μl, and the amine donor and acceptor concentrations were 25 mM each. All measurements were performed in triplicates at 40°C in a microplate reader at 600 nm (Synergy HT multimode microplate reader; BioTek) in continuous mode for a total time of 180 min. Specific activities (in U g^{-1} protein) were determined. One unit (U) of enzyme activity was defined as the amount of protein required to transform 1 μmol of substrate in 1 min under the assay conditions using a reaction product extinction coefficient (aldehyde 4 in reference 14) of 537 M^{-1} cm^{-1} at 600 nm, as determined experimentally. All values were corrected for nonenzymatic transformation (background rate).

Enzyme assays for determination of amine substrates, including enantiopure amines. Transaminase activity was assayed using benzaldehyde as the acceptor and structurally diverse amines in 96-well plates. Any other aldehyde, ketone, or keto acid may be used instead of benzaldehyde. Prior to the assay, a solution of 25 mM benzaldehyde as the acceptor and 1.0 mM PLP as the cofactor was first prepared in 100 mM K_2HPO_4 buffer, pH 7.5 (40 ml). Then, a stock solution of each amine was prepared in acetonitrile or buffer, depending on the solubility. Reaction assays, in 96-well microtiter plates, were started by

adding 2.5 μl of a protein solution (stock solution, 10.0 mg/ml in 100 mM K_2HPO_4 buffer, pH 7.5) to an assay mixture containing 185 μl of PLP-benzaldehyde solution and 12.5 μl of amine stock solution. The final volume of the assay mixture was 200 μl, and the amine donor and acceptor concentrations were 25 mM each. Reactions were allowed to proceed for 60 min at 40°C, during which time the amount of benzaldehyde remaining (not reacting with the amines) was determined every 5 min by adding 12.5 μl of a stock solution of the amine 2-(4-nitrophenyl)ethan-1-amine (400 mM in 100 mM K_2HPO_4 buffer, pH 7.5). After adding 2-(4-nitrophenyl)ethan-1-amine, the reaction was allowed to proceed for 10 min and absorbance due to the appearance of orange/red color was recorded continuously in a microplate reader at 600 nm (Synergy HT multimode microplate reader; BioTek). Lower absorbance values imply higher consumption of benzaldehyde and, thus, of the corresponding amines used as donors. Enzyme activity under the assay conditions was expressed as the amount of enzyme required to transform 1 μmol of substrate in 1 min under the assay conditions using a reaction product extinction coefficient (aldehyde 4 in reference 14) of 537 M^{-1} cm^{-1} at 600 nm. All values were corrected for nonenzymatic transformation (background rate) and for the results from a control reaction mixture containing benzaldehyde but not amines [no transfer reaction occurs, so all of the benzaldehyde reacts with 2-(4-nitrophenyl)ethan-1-amine].

Mass spectrometry. Conventional mass spectrometry analyses were performed on a hybrid quadrupole time-of-flight (Q-TOF) analyzer (QSTAR Pulsar I; AB Sciex, Framingham, MA, USA). Reaction samples were analyzed by direct infusion and ionized by electrospray ionization-mass spectrometry (ESI-MS) with methanol as the mobile phase in positive reflector mode. High-resolution mass spectrometry (HR-MS) analysis was carried out by flow injection analysis combined with electrospray ionization-mass spectrometry (FIA-ESI-MS) on an Agilent G6530A accurate-mass Q-TOF liquid chromatography-mass spectrometry (LC-MS) system (Agilent Technologies, Santa Clara, CA, USA). The sample was directly infused and ionized by ESI in negative reflector mode. Ionization was enhanced by JetStream technology, and the mobile phase was 99.9:0.1 (vol/vol) H_2O–formic acid. Data were processed with MassHunter Data Acquisition B.05.01 and MassHunter Qualitative Analysis B.07.00 software (Agilent Technologies).

GC analysis for determination of chiral selectivity. Enantioselectivity was evaluated by kinetic resolution of (*R*) and (*S*) amines. Prior to the kinetic resolution assay, a solution of 25 mM benzaldehyde as the acceptor and 1.0 mM PLP as a cofactor was first prepared in 100 mM K_2HPO_4 buffer, pH 7.5 (40 ml). A stock solution of a racemic mixture of (*R*)- and (*S*)-aminononane at a concentration of 400 mM each in acetonitrile was prepared. Reaction assays, in 5.0-ml Eppendorf tubes, were started by adding 25 μl of a protein solution (stock solution, 10.0 mg/ml in 100 mM K_2HPO_4 buffer, pH 7.5) to an assay mixture containing 1,850 μl of PLP-benzaldehyde solution and 125 μl of racemic (*R*)-/(*S*)-aminononane stock solution. The final volume of the assay mixture was 2,000 μl, and the concentrations of (*R*)- and (*S*)-aminononane and benzaldehyde were 25 mM each. Reactions were allowed to proceed at 40°C for 60 min. Next, the reaction mixture was filtered through an adsorptive, hydrophilic, 3K-nominal-molecular-weight-cutoff membrane (regenerated cellulose, Amicon) to remove the enzymes. Then, 10 μl of a stock solution of (*R*)-2-aminohexane was added as an internal standard to take into consideration biases due to the extraction procedure, as follows. To 0.2 ml of the reaction mixture, 0.2 ml ethyl acetate was added, and after vigorous vortexing, the solvent used for GC-MS analysis. The GC system (Agilent Technologies 7890A) consisted of an autosampler (Agilent Technologies 7693) and an inert MSD (mass selective detection) instrument with quadrupole (Agilent Technologies 5975). A total of 2 μl of the sample was injected through a CP-Chirasil-Dex CB GC column (25 m in length, 0.25-mm internal diameter, 0.25-μm film) (J&W GC columns; Agilent). The flow rate of the helium carrier gas, the split ratio, and the temperature gradient were optimized for each of the chiral mixes. After each injection, the column was cleaned up during 2 min at 200°C using a 1.5-ml/min flow rate. The detection of each chiral compound [(*R*)- and (*S*)-aminononane] was performed in single-ion-monitoring (SIM) mode in order to maximize the sensitivity. The elution order and the target ions were previously validated with a mixture of standards and the NIST 14 library. The semiquantification of (*R*)- and (*S*)-aminononanes was performed using MassHunter Qualitative Analysis software (B.06.00; Agilent), reporting the area for the individual peaks in arbitrary units, on the basis of which enantiomeric excess and conversion were calculated.

Stereochemistry was also calculated by following the asymmetric synthesis of (*R*)- and (*S*)-aminononane. Asymmetric synthesis assays were performed at 40°C for 180 min in 100 mM K_2HPO_4 buffer, pH 7.5, containing 1 mM PLP and 25 μg pure protein. The reaction mixture contained 25 mM 2-nonane as the acceptor and 25 mM 2-(4-nitrophenyl)ethan-1-amine as the amine donor. The conversion was measured by detection of the formed amines (*R*)-aminononane and (*S*)-aminononane by GC after extraction of the reaction products as described above. The percent-enantiomeric-excess values for products were analyzed by GC. Note that conversions were not optimized.

Enzyme assays for determinations of optimal parameters for activity. Temperatures between 25 and 70°C were tested to determine the conditions under which each protein displayed maximal activity in 100 mM K_2HPO_4 buffer, pH 7.5. Assays of activity in the presence of a water-miscible solvent were performed by adding methanol, acetonitrile, dimethyl sulfoxide, dimethyl acetamide, isopropanol, or acetone at concentrations from 5% to 50% (vol/vol). Benzaldehyde and 2-(4-nitrophenyl)ethan-1-amine were used as the acceptor and donor, respectively. The standard assay conditions to determine optimal temperatures and activities in the presence of solvents were as for determinations of acceptor substrates (see above).

Homology modeling and docking simulations. Homology models were developed using Prime software from Schrödinger. Prime uses BLAST (with the BLOSUM62 matrix) for homology search and alignment and refines the results using the Pfam database and pairwise alignment with ClustalW. Docking simulations of (*S*)-(+)-2-aminononane and (*R*)-(−)-2-aminononane with the structural models

created for each of the class III ω-TAs were carried out using the Protein Energy Landscape Exploration software, which offers one of the best modeling alternatives to map protein-ligand dynamics and induced fit, as described previously (48). The substrate was initially positioned in the active site with the nitrogen atom of the substrate toward the Lys catalytic base. The substrate conformation was set to be fully flexible in the docking simulations, whereas the protein conformation was not allowed to change.

Genetic enzymology analysis. Sequence similarity networks (SSNs) were generated by using the Enzyme Function Initiative's Enzyme Similarity Tool (EFI-EST) (51). All-vs-all BLAST was performed against the first 500 BLAST hits in UniProt (2018_06) for each query sequence (option A). A negative LogE value was applied for the initial network generation. The network consists of the nodes representing protein clusters with >60% sequence identity. The network was visualized in Cytoscape version 3.3.0 (52), using the organic layout in which the lengths of the edges correlate with the dissimilarity of the connected sequences represented by the nodes. The subgroupings in each major cluster were visualized by gradually increasing the stringency of the LogE filter of the networks. The final network was used to build genome neighborhood networks (GNNs) in EFI's Genome Neighborhood Network Tool (EFI-GNT). Initial default values of a 10-ORF window and 20% cooccurrence were chosen, although the values were eventually narrowed to a 5-ORF window and 20% cooccurrence. The GNT database uses the updated UniPro 2018_06 and ENA 136 versions. Accession numbers within relevant Pfam nodes were extracted and used for building a new SSN using option D. The published functional data were used to determine the consensus function and substrate preference of each subfamily of protein sequences.

Accession number(s). The sequences were named based on the code TR, which refers to transaminase, followed by an arbitrary number representing 1 of the 10 enzymes analyzed. Sequences encoding the enzymes designated TR_1 to TR_8 were deposited at the NCBI public database under accession numbers MF158200, MH588437, MF158202, MF158203, MF158204, MF158205, MF158206, and MF158207. Sequences encoding TR_9 and TR_{10} are available as part of the genome sequence of *P. oleovorans* (accession numbers NZ_NIUB01000001 and NZ_NIUB01000017).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/AEM.02404-18.

SUPPLEMENTAL FILE 1, PDF file, 3 MB.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation program (Blue Growth: Unlocking the Potential of Seas and Oceans) through Project INMARE under grant agreement no. 634486 and ERA-IB 5 METACAT. This work was further funded by grants no. PCIN-2014-107 (within the ERA NET IB2 grant no. ERA-IB-14-030—MetaCat) and PCIN-2017-078 (within the Marine Biotechnology ERA-NET [ERA-MBT], funded under the European Commission's Seventh Framework Program, 2013 to 2017, grant agreement no. 604814) and BIO2014-54494-R and BIO2017-85522-R from the Ministerio de Ciencia, Innovación y Universidades, formerly Ministerio de Economía, Industria y Competitividad. The present investigation was also funded by grant no. BB/M029085/1 from the UK Biotechnology and Biological Sciences Research Council (BBSRC). P.N.G. and R.B. acknowledge the support of the Supercomputing Wales project, which is partly funded by the European Regional Development Fund (ERDF) via the Welsh Government. O.V.G. and P.N.G. acknowledge the support of the Centre of Environmental Biotechnology Project funded by the European Regional Development Fund (ERDF) through the Welsh Government. We gratefully acknowledge financial support provided by the European Regional Development Fund (ERDF). C.C. thanks the Spanish Ministry of Economy, Industry and Competitiveness for a Ph.D. fellowship (grant no. BES-2015-073829).

We acknowledge Bayer AG for kindly providing some of the amines tested. We also acknowledge María J. Vicente at the Servicio Interdepartamental de Investigación (SIDI) from the Autonomous University of Madrid for the ESI-MS analyses.

C.C., N.K., J.N.-F., D.A., M.M.-M., A.B., and M.F. contributed to protein purification and characterization. C.C. and M.F. contributed three-dimensional modeling. C.G., T.N.C., H.T., O.V.G., M.M.Y., P.N.G., and K.-E.J. contributed to sample collection and library construction. R.B., A.G.-M., G.E.K.B., and M.F. performed sequence, phylogenetic, and genomic analyses. N.K., A.G.-M., A.B., and G.E.K.B. contributed to gene cloning, synthesis and expression. D.R. and C.B. carried out gas chromatography analyses. R.K. contributed suggestions for activity tests and provided substrates. M.F., K.-E.J., and P.N.G. conceived the work. M.F. wrote the initial draft of the manuscript.

REFERENCES

1. Steffen-Munsberg F, Vickers C, Kohls H, Land H, Mallin H, Nobili A, Skalden L, van den Bergh T, Joosten HJ, Berglund P, Höhne M, Bornscheuer UT. 2015. Bioinformatic analysis of a PLP-dependent enzyme superfamily suitable for biocatalytic applications. Biotechnol Adv 33:566–604. https://doi.org/10.1016/j.biotechadv.2014.12.012.
2. Fuchs M, Farnberger JE, Kroutil W. 2015. The industrial age of biocatalytic transamination. European J Org Chem 2015:6965–6982. https://doi.org/10.1002/ejoc.201500852.
3. Anton BP, Chang YC, Brown P, Choi HP, Faller LL, Guleria J, Hu Z, Klitgord N, Levy-Moonshine A, Maksad A, Mazumdar V, McGettrick M, Osmani L, Pokrzywa R, Rachlin J, Swaminathan R, Allen B, Housman G, Monahan C, Rochussen K, Tao K, Bhagwat AS, Brenner SE, Columbus L, de Crécy-Lagard V, Ferguson D, Fomenkov A, Gadda G, Morgan RD, Osterman AL, Rodionov DA, Rodionova IA, Rudd KE, Söll D, Spain J, Xu SY, Bateman A, Blumenthal RM, Bollinger JM, Chang WS, Ferrer M, Friedberg I, Galperin MY, Gobeill J, Haft D, Hunt J, Karp P, Klimke W, Krebs C, Macelis D, et al. 2013. The COMBREX project: design methodology and initial results. PLoS Biol 11:e1001638. https://doi.org/10.1371/journal.pbio.1001638.
4. Colin Y, Kintses B, Gielen F, Miton CM, Fischer G, Mohamed MF, Hyvönen M, Morgavi DP, Janssen DB, Hollfelder F. 2015. Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. Nat Commun 6:10008. https://doi.org/10.1038/ncomms10008.
5. Ferrer M, Martínez-Martínez M, Bargiela R, Streit WR, Golyshina OV, Golyshin PN. 2016. Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. Microb Biotechnol 9:22–34. https://doi.org/10.1111/1751-7915.12309.
6. Martínez-Martínez M, Bargiela R, Coscolín C, Navarro-Fernández J, Golyshin PN, Ferrer M. 2016. Functionalization and modification of hydrocarbon-like molecules guided by metagenomics: enzymes most requested at the industrial scale for chemical synthesis as study cases, p 1–26. *In* Lee SY (ed), Consequences of microbial interactions with hydrocarbons oils and lipids: production of fuels and chemicals. Springer International Publishing AG, Cham, Switzerland.
7. Chen C, Huang H, Wu CH. 2017. Protein bioinformatics databases and resources. Methods Mol Biol 1558:3–39. https://doi.org/10.1007/978-1-4939-6783-4_1.
8. Ferrandi EE, Previdi A, Bassanini I, Riva S, Peng X, Monti D. 2017. Novel thermostable amine transferases from hot spring metagenomes. Appl Microbiol Biotechnol 101:4963–4979. https://doi.org/10.1007/s00253-017-8228-2.
9. Han SW, Kim J, Cho H-S, Shin J-S. 2017. Active site engineering of ω-transaminase guided by docking orientation analysis and virtual activity screening. ACS Catal 7:3752–3762. https://doi.org/10.1021/acscatal.6b03242.
10. Han SW, Park ES, Dong JY, Shin JS. 2015. Mechanism-guided engineering of ω-transaminase to accelerate reductive amination of ketones. Adv Synth Catal 357:1732–1740. https://doi.org/10.1002/adsc.201500211.
11. Svedendahl M, Branneby C, Lindberg L, Berglund P. 2010. Reversed enantiopreference of an ω-transaminase by a single-point mutation. ChemCatChem 2:976–980. https://doi.org/10.1002/cctc.201000107.
12. Steffen-Munsberg F, Vickers C, Thontowi A, Schätzle S, Tumlirsch T, Svedendahl Humble M, Land H, Berglund P, Bornscheuer UT, Höhne M. 2013. Connecting unexplored protein crystal structures to enzymatic function. ChemCatChem 5:150–153. https://doi.org/10.1002/cctc.201200544.
13. Skalden L, Peters C, Dickerhoff J, Nobili A, Joosten HJ, Weisz K, Höhne M, Bornscheuer UT. 2015. Two subtle amino acid changes in a transaminase substantially enhance or invert enantiopreference in cascade syntheses. Chembiochem 16:1041–1045. https://doi.org/10.1002/cbic.201500074.
14. Baud D, Ladkau N, Moody TS, Ward JM, Hailes HC. 2015. A rapid sensitive colorimetric assay for the high-throughput screening of transaminases in liquid or solid-phase. Chem Commun (Camb) 51:17225–17228. https://doi.org/10.1039/c5cc06817g.
15. Green AP, Turner NJ, O'Reilly E. 2014. Chiral amine synthesis using ω-transaminases: an amine donor that displaces equilibria and enables high-throughput screening. Angew Chem Int Ed 53:10714–10717. https://doi.org/10.1002/anie.201406571.
16. Martínez-Martínez M, Coscolín C, Santiago G, Chow J, Stogios PJ, Bargiela R, Gertler C, Navarro-Fernández J, Bollinger A, Thies S, Méndez-García C, Popovic A, Brown G, Chernikova TN, García-Moyano A, Bjerga GEK, Pérez-García P, Hai T, Del Pozo MV, Stokke R, Steen IH, Cui H, Xu X, Nocek BP, Alcaide M, Distaso M, Mesa V, Peláez AI, Sánchez J, Buchholz PCF, Pleiss J, Fernández-Guerra A, Glöckner FO, Golyshina OV, Yakimov MM, Savchenko A, Jaeger KE, Yakunin AF, Streit WR, Golyshin PN, Guallar V, Ferrer M, The INMARE Consortium. 2018. Determinants and prediction of esterase substrate promiscuity patterns. ACS Chem Biol 13:225–234. https://doi.org/10.1021/acschembio.7b00996.
17. Bargiela R, Gertler C, Magagnini M, Mapelli F, Chen J, Daffonchi D, Golyshin PN, Ferrer M. 2015. Degradation network reconstruction in uric acid and ammonium amendments in oil-degrading marine microcosms guided by metagenomic data. Front Microbiol 6:1270. https://doi.org/10.3389/fmicb.2015.01270.
18. Bargiela R, Mapelli F, Rojo D, Chouaia B, Tornés J, Borin S, Richter M, Del Pozo MV, Cappello S, Gertler C, Genovese M, Denaro R, Martínez-Martínez M, Fodelianakis S, Amer RA, Bigazzi D, Han X, Chen J, Chernikova TN, Golyshina OV, Mahjoubi M, Jaouanil A, Benzha F, Magagnini M, Hussein E, Al-Horani F, Cherif A, Blaghen M, Abdel-Fattah YR, Kalogerakis N, Barbas C, Malkawi HI, Golyshin PN, Yakimov MM, Daffonchio D, Ferrer M. 2015. Bacterial population and biodegradation potential in chronically crude oil-contaminated marine sediments are strongly linked to temperature. Sci Rep 5:11651. https://doi.org/10.1038/srep11651.
19. Peña-García C, Martínez-Martínez M, Reyes-Duarte D, Ferrer M. 2016. High throughput screening of esterases lipases and phospholipases in mutant and metagenomic libraries: a review. CCHTS 19:605–615. https://doi.org/10.2174/1386207319666151110123927.
20. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389.
21. Rudat J, Brucher BR, Syldatk C. 2012. Transaminases for the synthesis of enantiopure beta-amino acids. AMB Express 2:11. https://doi.org/10.1186/2191-0855-2-11.
22. Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454.
23. Ménigaud S, Mallet L, Picord G, Churlaud C, Borrel A, Deschavanne P. 2012. GOHTAM: a website for 'Genomic Origin of Horizontal Transfers, Alignment and Metagenomics'. Bioinformatics 28:1270–1271. https://doi.org/10.1093/bioinformatics/bts118.
24. Bjerga GEK, Arsın H, Larsen Ø, Puntervoll P, Kleivdal HT. 2016. A rapid solubility-optimized screening procedure for recombinant subtilisins in *E. coli*. J Biotechnol 222:38–46. https://doi.org/10.1016/j.jbiotec.2016.02.009.
25. Katzke N, Knapp A, Loeschcke A, Drepper T, Jaeger K-E. 2017. Novel tools for the functional expression of metagenomic DNA. Methods Mol Biol 1539:159–196. https://doi.org/10.1007/978-1-4939-6691-2_10.
26. Pavlidis IV, Weiß MS, Genz M, Spurr P, Hanlon SP, Wirz B, Iding H, Bornscheuer UT. 2016. Identification of (S)-selective transaminases for the asymmetric synthesis of bulky chiral amines. Nat Chem 8:1076–1082. https://doi.org/10.1038/nchem.2578.
27. Han SW, Park ES, Dong JY, Shin JS. 2015. Active-site engineering of ω-transaminase for production of unnatural amino acids carrying a side chain bulkier than an ethyl substituent. Appl Environ Microbiol 81:6994–7002. https://doi.org/10.1128/AEM.01533-15.
28. Höhne M, Schätzle S, Jochens H, Robins K, Bornscheuer UT. 2010. Rational assignment of key motifs for function guides in silico enzyme identification. Nat Chem Biol 6:807–813. https://doi.org/10.1038/nchembio.447.
29. Weiß MS, Pavlidis IV, Spurr P, Hanlon SP, Wirz B, Iding H, Bornscheuer UT. 2017. Amine transaminase engineering for spatially bulky substrate acceptance. Chembiochem 18:1022–1026. https://doi.org/10.1002/cbic.201700033.
30. Park ES, Kim M, Shin JS. 2012. Molecular determinants for substrate selectivity of ω-transaminases. Appl Microbiol Biotechnol 93:2425–2435. https://doi.org/10.1007/s00253-011-3584-9.
31. Wu HL, Zhang JD, Zhang CF, Fan XJ, Chang HH, Wei WL. 2017. Characterization of four new distinct ω-transaminases from *Pseudomonas putida* NBRC 14164 for kinetic resolution of racemic amines and amino alcohols. Appl Biochem Biotechnol 181:972–985. https://doi.org/10.1007/s12010-016-2263-9.
32. Mallin H, Höhne M, Bornscheuer UT. 2014. Immobilization of (R - and

(S -amine transaminases on chitosan support and their application for amine synthesis using isopropylamine as donor. J Biotechnol 191:32–37. https://doi.org/10.1016/j.jbiotec.2014.05.015.
33. Sayer C, Martinez-Torres RJ, Richter N, Isupov MN, Hailes HC, Littlechild JA, Ward JM. 2014. The substrate specificity enantioselectivity and structure of the (R -selective amine:pyruvate transaminase from *Nectria haematococca*. FEBS J 28:2240–2253. https://doi.org/10.1111/febs.12778.
34. Gärtner A, Wiese J, Imhoff JF. 2008. *Amphritea atlantica* gen. nov., sp. nov., a gammaproteobacterium from the Logatchev hydrothermal vent field. Int J Syst Evol Microbiol 58:34–39. https://doi.org/10.1099/ijs.0.65234-0.
35. Khaleque HN, Ramsay JP, Murphy RJ, Kaksonen AH, Boxall NJ, Watkin EL. 2017. Draft genome sequence of the acidophilic, halotolerant, and iron/sulfur-oxidizing Acidihalobacter prosperus DSM 14174 (strain V6). Genome Announc 5:e01469-16. https://doi.org/10.1128/genomeA.01469-16.
36. Huang J, Xie DF, Feng Y. 2017. Engineering thermostable (R -selective amine transaminase from *Aspergillus terreus* through in silico design employing B-factor and folding free energy calculations. Biochem Biophys Res Commun 483:397–402. https://doi.org/10.1016/j.bbrc.2016.12.131.
37. Iglesias C, Panizza P, Rodriguez Giordano S. 2017. Identification expression and characterization of an R-ω-transaminase from *Capronia semiimmersa*. Appl Microbiol Biotechnol 101:5677–5687. https://doi.org/10.1007/s00253-017-8309-2.
38. Jiang J, Chen X, Zhang D, Wu Q, Zhu D. 2015. Characterization of (R -selective amine transaminases identified by in silico motif sequence blast. Appl Microbiol Biotechnol 99:2613–2621. https://doi.org/10.1007/s00253-014-6056-1.
39. Łyskowski A, Gruber C, Steinkellner G, Schürmann M, Schwab H, Gruber K, Steiner K. 2014. Crystal structure of an (R -selective ω-transaminase from *Aspergillus terreus*. PLoS One 9:e87350. https://doi.org/10.1371/journal.pone.0087350.
40. Calvelage S, Dörr M, Höhne M, Bornscheuer UT. 2017. A systematic analysis of the substrate scope of (S - and (R -selective amine transaminases. Adv Synth Catal 359:4235–4243. https://doi.org/10.1002/adsc.201701079.
41. Wortham BW, Patel CN, Oliveira MA. 2007. Polyamines in bacteria: pleiotropic effects yet specific mechanisms. Adv Exp Med Biol 603:106–115. https://doi.org/10.1007/978-0-387-72124-8_9.
42. Popovic A, Hai T, Tchigvintsev A, Hajighasemi M, Nocek B, Khusnutdinova AN, Brown G, Glinos J, Flick R, Skarina T, Chernikova TN, Yim V, Brüls T, Paslier DL, Yakimov MM, Joachimiak A, Ferrer M, Golyshina OV, Savchenko A, Golyshin PN, Yakunin AF. 2017. Activity screening of environmental metagenomic libraries reveals novel carboxylesterase families. Sci Rep 7:44103. https://doi.org/10.1038/srep44103.
43. Nacke H, Will C, Herzog S, Nowka B, Engelhaupt M, Daniel R. 2011. Identification of novel lipolytic genes and gene families by screening of metagenomic libraries derived from soil samples of the German biodiversity exploratories. FEMS Microbiol Ecol 78:188–201. https://doi.org/10.1111/j.1574-6941.2011.01088.x.
44. Troeschel SC, Drepper T, Leggewie C, Streit WR, Jaeger KE. 2010. Novel tools for the functional expression of metagenomic DNA. Methods Mol Biol 668:117–139. https://doi.org/10.1007/978-1-60761-823-2_8.
45. Placido A, Hai T, Ferrer M, Chernikova TN, Distaso M, Armstrong D, Yakunin AF, Toshchakov SV, Yakimov MM, Kublanov IV, Golyshina OV, Pesole G, Ceci LR, Golyshin PN. 2015. Diversity of hydrolases from hydrothermal vent sediments of the Levante Bay Vulcano Island (Aeolian archipelago) identified by activity-based metagenomics and biochemical characterization of new esterases and an arabinopyranosidase. Appl Microbiol Biotechnol 99:10031–10046. https://doi.org/10.1007/s00253-015-6873-x.
46. Poehlein A, Daniel R, Thürmer A, Bollinger A, Thies S, Katzke N, Jaeger K-E. 2017. First insights into the genome sequence of *Pseudomonas oleovorans* DSM1045. Genome Announc 5:e00774-17. https://doi.org/10.1128/genomeA.00774-17.
47. Geertsma ER, Dutzler R. 2011. A versatile and efficient high-throughput cloning tool for structural biology. Biochemistry 50:3272–3278. https://doi.org/10.1021/bi200178z.
48. Santiago G, Martínez-Martínez M, Alonso S, Bargiela R, Coscolín C, Golyshin PN, Guallar V, Ferrer M. 2018. Rational engineering of multiple active sites in an ester hydrolase. Biochemistry 57:2245–2255. https://doi.org/10.1021/acs.biochem.8b00274.
49. Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685. https://doi.org/10.1038/227680a0.
50. Bradford MM. 1976. A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254. https://doi.org/10.1016/0003-2697(76)90527-3.
51. Gerlt JA, Allen KN, Almo SC, Armstrong RN, Babbitt PC, Cronan JE, Dunaway-Mariano D, Imker HJ, Jacobson MP, Minor W, Poulter CD, Raushel FM, Sali A, Shoichet BK, Sweedler JV. 2011. The Enzyme Function Initiative. Biochemistry 50:9950–9962. https://doi.org/10.1021/bi201312u.
52. Shannon P, Markiel A, Ozier O, Balinaga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoescape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504. https://doi.org/10.1101/gr.1239303.

Chapter 4: Controlled manipulation of enzyme specificity through immobilization-induced flexibility constraints.

In this work, we continue with the idea of turning promiscuous but not enantio-selective enzymes into promiscuous and enantio-selective enzymes. This phenotype, which is widely appreciated in industry is, however, uncommon in nature, as discussed in Chapters 1 and 2. As we know that the flexibility of a protein may be also a structural determinant for its catalytic properties we wanted to explore the concept of the so-called supramolecular engineering by analyzing the consequences of changing the flexibility of the protein scaffold in the specificity and selectivity features. This was done by applying multiple immobilization techniques to an ester-hydrolase.

The idea was to start with enzymes with proven substrate promiscuity and use these as a lead scaffold to produce better variants in terms of expanded substrate specificity and selectivity. The EH1 ester-hydrolase was selected for this purpose as being the top substrate promiscuous ester-hydrolase in our collection, as it is not enantio-selective. In this Chapter, we selected both a mesoporous material and micro-particles for immobilizing the previously mentioned ester-hydrolase. We first ensured that the size of the pore of the mesoporous material was significantly bigger than that of EH1 so that the enzyme can enter the pore and remain inside the material and the substrates can freely diffuse in and out of the support. In this case, as no covalent interactions exist, the free movement of the enzyme was guaranteed. In a second step, we designed a mesoporous material with pore sizes close to that of the enzyme so that the protein can enter but the movement is constrained. Finally, and in a similar way, the enzyme was immobilized on the surface of two similar micro-particles but through linkers of different natures and lengths, so that the covalent linking of the enzyme can be associated with differences in flexibility/rigidity. A total of 4 bio-catalysts were prepared, which when tested with the 96 esters used in Chapter 1 allowed evaluating the effect of the different immobilization techniques both in the substrate profiles and the enantio-selectivity. The major outcome was that a promiscuous but not enantio-selective enzyme can be transformed into a promiscuous enzyme with stringent stereo-specificity when the flexibility of the enzyme is constrained. Also, depending on the flexibility constraints one can create biocatalysts with the capacity to convert different chiral molecules with stringent stereo-specificity in all cases. Together, by diminishing or differentially influencing the level of flexibility of an enzyme during immobilization, one can differentially influence the access and positions of different substrates, including chiral ones, in the active site.

The results of this Chapter may open future possibilities to study the level of flexibility/rigidity that different immobilization tools produced in enzymes and their catalytic consequences. In addition to that, these influences may produce additional benefits to the immobilization process itself, which can be beneficial for protecting the enzyme against organic solvents, temperatures or other extreme conditions, typically used in the industrial sector. The economic saving by re-utilizing the biocatalysts is an important characteristic that makes immobilized enzymes very suitable for their application in biotech companies.

Applied Catalysis A, General 565 (2018) 59–67

Contents lists available at ScienceDirect

Applied Catalysis A, General

journal homepage: www.elsevier.com/locate/apcata

Controlled manipulation of enzyme specificity through immobilization-induced flexibility constraints

Cristina Coscolín[a], Ana Beloqui[b], Mónica Martínez-Martínez[a], Rafael Bargiela[a,c], Gerard Santiago[d], Rosa M. Blanco[a], Guillaume Delaittre[e,f], Carlos Márquez-Álvarez[a,*], Manuel Ferrer[a,*]

[a] Institute of Catalysis, Consejo Superior de Investigaciones Científicas, 28049 Madrid, Spain
[b] CICnanoGUNE, 20018 San Sebastián, Spain
[c] School of Chemistry, Bangor University, LL57 2UW Bangor, UK
[d] Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain
[e] Karlsruhe Institute of Technology, Institute of Toxicology and Genetics, 76344 Eggenstein-Leopoldshafen, Germany
[f] Karlsruhe Institute of Technology, Institute for Chemical Technology and Polymer Chemistry, 76131 Karlsruhe, Germany

ARTICLE INFO

Keywords:
Biocatalyst
Esterase
Immobilization
Mesoporous
Microparticles
Specificity

ABSTRACT

It is thought that during immobilization enzymes, as dynamic biomolecules, may become distorted and this may alter their catalytic properties. However, the effects of different immobilization strategies on enzyme rigidity or flexibility and their consequences in specificity and stereochemistry at large scale has not been yet clearly evaluated and understood. This was here investigated by using as model an ester hydrolase, isolated from a bacterium inhabiting a karstic lake, with broad substrate spectrum (72 esters being converted; 61.5 U mg^{-1} for glyceryl tripropionate) but initially non-enantiospecific. We found that the enzyme (7 nm × 4.4 nm × 4.2 nm) could be efficiently ionic exchanged inside the pores (9.3 nm under dry conditions) of amino-functionalized ordered mesoporous material (NH_2-SBA-15), achieving a protein load of 48 mg g^{-1}, and a specific activity of 4.5 ± 0.1 U mg^{-1}. When the enzyme was site-directed immobilized through His interaction with an immobilized cationon the surface of two types of magnetic micro-particles through hexahistidine-tags, protein loads up to 10.2 µg g^{-1} and specific activities of up to 29.9 ± 0.3 U mg^{-1}, were obtained. We found that ionically exchanged enzyme inside pores of NH_2-SBA-15 drastically narrowed the substrate range (17 esters), to an extent much higher than ionically exchanged enzyme on the surface of magnetic micro-particles (up to 61 esters). This is attributed to differences in surface chemistry, particle size, and substrate accessibility to the active site tunnel. Our results also suggested, for the first time, that immobilization of enzymes in pores of similar size may alter the enzyme structures and produce enzyme active centers with different configuration which promote stereochemical conversions in a manner different to those arising from surface immobilization, where the strength of the ionic exchange also has an influence. This was shown by demonstrating that when the enzyme was introduced inside pores with a diameter (under dry conditions) slightly higher than that of the enzyme crystal structure a biocatalyst enantiospecific for ethyl (*R*)-4-chloro-3-hydroxybutyrate was produced, a feature not found when using wider pores. By contrast, immobilization on the surface of ferromagnetic microparticles produced selective biocatalysts for methyl (*S*)-(+)-mandelate or methyl (*S*)-lactate depending on the functionalization. This study illustrates the benefits of extensive analysis of the substrate spectra to better understand the effects of different immobilization strategies on enzyme flexibility/rigidity, as well as substrate specificity and stereochemistry. Our results will help to design tunable materials and interfaces for a controlled manipulation of specificity and to transform non-enantiospecific enzymes into stereo-chemically substrate promiscuous biocatalysts capable of converting multiple chiral molecules.

* Corresponding authors.
E-mail addresses: c.marquez@icp.csic.es (C. Márquez-Álvarez), mferrer@icp.csic.es (M. Ferrer).

https://doi.org/10.1016/j.apcata.2018.08.003
Received 8 June 2018; Received in revised form 2 August 2018; Accepted 4 August 2018
Available online 06 August 2018

1. Introduction

Enzymes are delicate organic catalysts that need to be stabilized to survive a range of challenging conditions typically used in industrial processes. Multiple immobilization methods and materials have been successfully employed to generate stabilized biocatalysts [1–4], but the most efficient immobilization protocol and materials may be selected aiming at a balance between activity, specificity, stability, and costs [5]. Besides most used carriers such as (epoxy)acrylic resins and agarose and widely used nanoparticles such as chitosan and nanoflowers [6], outstanding recent examples for enzyme immobilization and stabilization included biomimetic silica supports [7], mesostructured onion-like silica [8], hybrid macroporous foams synthesized via an integrative chemistry synthetic pathway [9], superparamagnetic silica/iron oxide nanocomposites with mesostructured porosity [10], maghemite (γ-Fe_2O_3) nanoparticles functionalized with a reactive multifunctional polymer [11], dendronized polymer and mesoporous silica nanoparticles [12], carbon nanotubes and polymers [13], epoxy-activated carriers [14], borosilicate [15], hematite nanoparticles [16], phyto-inspired silica nanowires [17], and materials with bioinspired coatings [18], to cite some. These studies exemplify the large interest in designing new and more efficient carriers for enzyme immobilization, the comparative analysis of which at large, in relation to their effect in enzyme substrate specificity (including enantiospecificity), may deserve further interest.

A major advance in the last decade has been the development of enzyme encapsulation in ordered mesoporous materials, because of their high specific surface area and pore volume, their highly uniform and tunable pore sizes, and the possibility to create micropores interconnecting mesopores and to control their morphology; they are also thermally, mechanically and chemically stable [5,19–23]. Various proteins have been successfully immobilized on mesoporous materials, including, non-catalytic proteins, oxidoreductases, hydrolases, lyases and isomerases [5,24–41]. Enzymes immobilized in confined macromolecular environments where the surface and volume of the confined environment can be controlled, are suggested to constitute optimal enzymatic nanoreactors [19,42,43]. Their surface can be grafted with different functionalities [23,17,18], a property particularly interesting to allow, for example, the ionic exchange enzyme encapsulation in amino-functionalized ordered mesoporous materials. This development led to an increased enzyme loading (up to 187 mg g^{-1} of catalyst) by simply adjusting synthetic conditions of the siliceous material for the obtaining of pores large enough to permit enzyme entrance and diffusion through the pore channels [17,18]. As example, mesoporous materials with pore diameters as low as 3–10 nm and also tailored between 3–30 nm have been synthesized [for examples see ref. 19,23,25,28,29,31], and enzymes with sizes such as 6.1 nm × 5.0 nm × 4.9 nm [20] have been immobilized in the pores. In these ordered structures the enzyme is immobilized by electrostatic interactions with limited diffusional restrictions and low enzyme modification, typically occurring after covalent binding or crosslinking [44]. This ion exchange of the enzyme on the support, was demonstrated to minimize leaching and maximize activity and stability of the biocatalyst [17,18].

Compared to free enzyme in solution, immobilization on a surface often hinders the free movement of the enzyme, although it is believed that the enzyme is positioned in an environment where the inherent flexibility is high [45]. The confinement in the pores of mesoporous materials is also thought to hinder the free movement of enzyme, but to what extent the differences in movement and flexibility affects enzyme properties remains to be elucidated. In this direction, it is only supposed that larger substrates may be preferably hydrolyzed by enzymes immobilized on the external surface or at the entrance of the channels of mesoporous materials because of substrate diffusional and enzyme orientation issues, as it was demonstrated by modelling predictions [46] and also using a lipase when tested with small (tributyrin) and large (triolein) substrates [47]. However, a general overview about how enzyme immobilization inside pores affects substrate specificity compared to the free enzyme and surface-immobilized enzyme remains to be established. Indeed, diverse surface immobilization methods have been shown to preserve activity level and, most importantly, to even create enantiospecific enzymes such as lipases [for example see ref. 48–50]. These, and other studies, have been performed using a restricted set of molecules and a comparative analysis at large is lacking.

Here, we examine the substrate range and enantiospecificity of a target enzyme by using as supports, an amino-functionalized ordered mesoporous material (NH_2-SBA-15) allowing non-covalent enzyme immobilization inside pores and two types of magnetic microparticles for non-covalent surface immobilization. Particularly, agarose-coated ferromagnetic core-shell microparticles with a nitrilotetracetic acid (NTA) tetradentate ligand, and polyvinyl alcohol microparticles embedded with magnetite and grafted with an iminodiacetic acid (IDA) tridentate ligand. Both particles were Ni^{2+}-activated and thus can be used for purification and immobilization given the affinity for hexahistidine-tags. As model enzyme, a serine ester hydrolase isolated from the metagenomic DNA of microbial communities inhabiting a karstic lake [51], referred to as EH1, with a typical α/β hydrolase fold as a model. Its structure was recently solved (PDB code 5JD4) [51,52], and in a recent study it was identified as one of the ester hydrolases with broader substrate spectrum among a total of 147 when tested with a set of 96 chemically and structurally different esters [51]. The active site (catalytic triad: Ser161, Asp256 and His286; oxyanion hole: Gly88, Gly89 and Gly90) supports the hydrolysis of a broad range of 72 esters, with vinyl butyrate (200.7 ± 0.4 U mg^{-1}) and phenyl propionate (198.7 ± 0.9 U mg^{-1}) serving as the best substrates (Table S1). Being highly promiscuous in terms of substrate scope, EH1 is not enantiospecific; thus the apparent enantiospecificity (E_{app}) factor calculated as the ratio of specific activities for 14 chiral esters when pure stereoisomers were tested separately [53] was below 25, a value above which ester hydrolases begin to have practical value [54]. Based on these considerations, EH1 may be then considered as an ester hydrolase with prominent substrate promiscuity but with limited practical use due to the low stereospecificity. Because substrate promiscuity and enantiospecificity are two appreciated properties when combined in a single biocatalyst, which is rare in nature within esterases [54], we evaluated whether both properties can be tailored by employing three immobilization strategies.

To the best of our knowledge, detailed analysis of the substrate spectrum and interpretation of the data provides, for the first time, new insights into the contribution of chemical, physical, and flexibility constrains to the catalytic capacity of biocatalysts prepared by these immobilization methods. We would like to highlight that the effect of immobilization on enzyme flexibility has been previously studied, for example, by increasing the number of bonds by which the enzyme is linked to the surface of carriers [48–50]. It is though that this increases enzyme rigidity and in turn has a consequence in promoting enzyme stability and activity when tested over a specific set of substrates [48–50]. The novelty of this study is that through an analysis of the substrate specificity with a very broad spectrum of molecules, and 3 different immobilization techniques, we were able to demonstrate that each polymeric carrier induced different flexibility and rigidity constraints to the enzyme with distinct effects on specificity and, most importantly, enantiospecificity, which can be also controlled.

2. Experimental section

2.1. Chemicals and EH1 protein source

All chemicals used for enzymatic tests were of the purest grade available and were purchased from sources described elsewhere [51]. The isolation of the enzyme EH1, available in the expression plasmid pET46 Ek/LIC and *Escherichia coli* BL21 as a host, was reported previously [51,52].

2.2. Gene expression and protein purification

Expression of the gene encoding the N-terminally hexahistidine-tagged enzyme and purification were performed as previously described [51,52]. Purity was assessed as > 98% using SDS-PAGE analysis in a Mini PROTEAN electrophoresis system (Bio-Rad) [55]. The protein concentration was determined according to Bradford assay with bovine serum albumin as a standard [56].

2.3. Synthesis and characterization of the mesostructured silica support and EH1 immobilization

Mesoporous silica with a periodic arrangement of uniform-diameter pore channels (SBA-15 mesostructure) was synthesized according to the method reported by Linton and Alfredsson [57]. A 2.5 wt% aqueous solution of Pluronic PE-10400 (BASF) surfactant was prepared by carefully dissolving this compound in a 1.6 M HCl solution, at 35 °C, in a closed polypropylene container kept under magnetic stirring for 2 h. Tetramethyl orthosilicate (TMOS, 98%, Aldrich) was added to this solution (3.8 g TMOS in 100 g of surfactant solution) and the mixture kept at 55 °C for 20 h under magnetic stirring. The white solid product formed was aged in the mother liquor at 75 °C for 24 h without stirring. Finally, the solid was filtered off, washed with absolute ethanol (Panreac) and calcined in air in a muffle furnace at 550 °C for 6 h using a 2 °C min^{-1} heating ramp. The white powder obtained is labelled "SBA-15".

Powder X-ray diffraction (XRD) pattern of SBA-15 silica (Fig. 1) were recorded at low angle (0.4-6°, step size 0.0167°) with a PANalytical X'Pert Pro diffractometer using CuKα radiation with a nickel filter. The pattern confirms that the pore channels arrangement exhibits p6mm hexagonal symmetry. It shows an intense peak at a 2θ angle of 0.9° corresponding to the d_{100} spacing and two weak reflections at 2θ close to 1.5° and 1.7°, assigned to the d_{110} and d_{200} spacing, respectively. From these data, the calculated unit-cell lattice parameter a_0 is 11.3 nm. Note that the unit-cell lattice parameter a_0 corresponds to the distance between the centers of two neighboring cylindrical pore channels (that is, the sum of the pore diameter and of the width of the silica pore wall that separates two channels).

Scanning electron microscopy (SEM) images were recorded using a FEI Nova NanoSEM 230 FE-SEM microscope equipped with vCD detector. These images show (Fig. 2, left) the plate-like shape of silica particles. Platelets were several microns wide and their thickness was close to 200 nm. Transmission electron microscopy (TEM) images (obtained with a JEOL JEM-2100 electron microscope operated at 200 kV) showed (Fig. 2, right) that the mesopore channels exhibit a long-range 2D hexagonal arrangement, in agreement with XRD data, and that these channels run parallel to the short axis of the platelets.

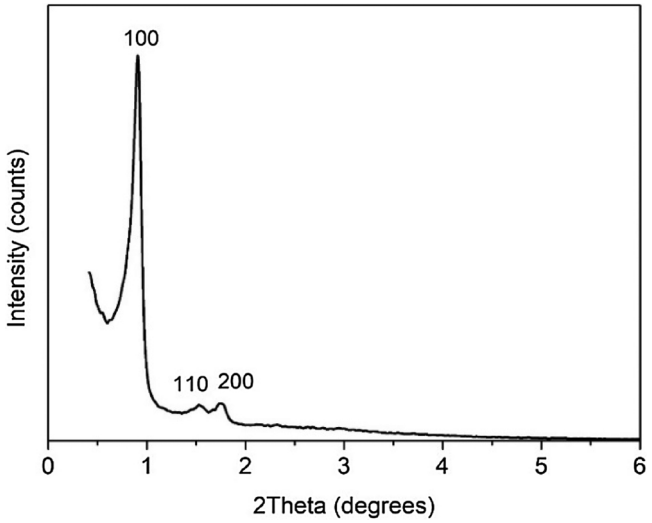


Fig. 1. XRD pattern of SBA-15 silica.

Nitrogen adsorption-desorption isotherm of SBA-15 silica (Fig. 3) was acquired at 77 K with a Micromeritics ASAP 2420 sorptometer. Prior to analysis, the sample was degassed at 623 K for 16 h. The isotherm obtained is of the type IVa according to 2015 IUPAC classification [58]. This type of isotherm shows a steep increase of the adsorbed volume (that takes place at a relative pressure around 0.7 in the isotherm plotted in Fig. 3), due to capillary condensation in mesopores, together with a hysteresis loop that is consistent with the presence of cylindrical mesopores. Pore size distribution (Fig. 3, inset) was obtained from the adsorption data using the methods of non-local density functional theory (NLDFT) with the kernel function for oxide materials with cylindrical pores [59,60]. The narrow distribution obtained indicates that the dried mesoporous silica possesses pore channels with a highly uniform diameter close to 9.3 nm. DFT calculations also indicated that mesopores account for a total volume of 0.44 $cm^3 g^{-1}$ and that the sample exhibits some micro-porosity (as expected for SBA-15 silica) accounting for an additional volume of 0.1 $cm^3 g^{-1}$.

The SBA-15 silica sample was functionalized with aminopropyl groups by grafting with an alkoxysilane. The silica powder (0.833 g) was placed inside a round bottom flask and dried overnight under vacuum (ca. 10^{-2} mbar) at 80 °C. Then, 50 mL of toluene (anhydrous, 99.8%, Sigma-Aldrich) were added to the flask and the solid kept in suspension with magnetic stirring under dry nitrogen atmosphere. To this suspension, 1.0 mL of 3-aminopropyl)triethoxysilane (APTES, 99%, Aldrich) was added (equivalent to 5 mmol alkoxysilane per gram of silica) and the mixture kept under reflux for 24 h. The solid was finally filtered off and washed with dry toluene and acetone. The content of amino groups in the functionalized sample was determined to be 1.0 mmol g^{-1}, as measured by CHN elemental analysis using a LECO CHNS-932 equipment. This sample, labelled "NH_2-SBA-15", was used as support for immobilization of EH1.

Briefly, a total of 0.96 mL of esterase solution (5 mg mL^{-1}) was added to 10.04 mL of 40 mM HEPES buffer pH 7.0. Once dissolved, 100 mg NH_2-SBA-15 silica were added to 10 mL of this solution. Catalytic activity of the protein solution before immobilization was measured using *p*-nitrophenyl acetate (*p*NP-acetate) as substrate as described previously [52]. Immobilization was allowed to proceed at different intervals after which the catalytic activity was evaluated in the supernatant and on the mesoporous silica (prior re-suspension in HEPES (40 mM, pH 7.0)) previously separated by centrifugation (13000 rpm, room temperature, 1.5 min), and compared to the activity of the initial esterase solution. The immobilization proceeded until the activity of the supernatant was constant, which indicated full enzyme immobilization. Then, the suspension was filtered using sintered glass funnel, washed with HEPES (40 mM, pH 7.0), and vacuum-dried. The specific activity of the immobilized enzyme was evaluated by quantifying the total units of esterase activity (measured using *p*NP-acetate as substrate) at the beginning of the experiment and the total units of the immobilized biocatalyst (tested using 9.4 mg of the biocatalyst re-suspended in 1 mL HEPES buffer, pH 7.0), as reported previously [52]. Under these conditions, after 48 h, only 4% of the initial activity remained in the supernatant and the enzyme loading obtained was of 48 mg esterase per gram of NH_2-SBA-15 (or 96% total protein being immobilized) after calculation of protein concentrations according to Bradford assay with bovine serum albumin as the standard [56].

Fig. 4 summarizes the protocol used for the synthesis of the mesostructured silica support and EH1 immobilization, performed at pH 7.0, at which the amino groups of the support are positively charged and the net charge of the enzyme is negative. Additional comments on enzyme immobilization kinetics is provided in the Supporting Information.

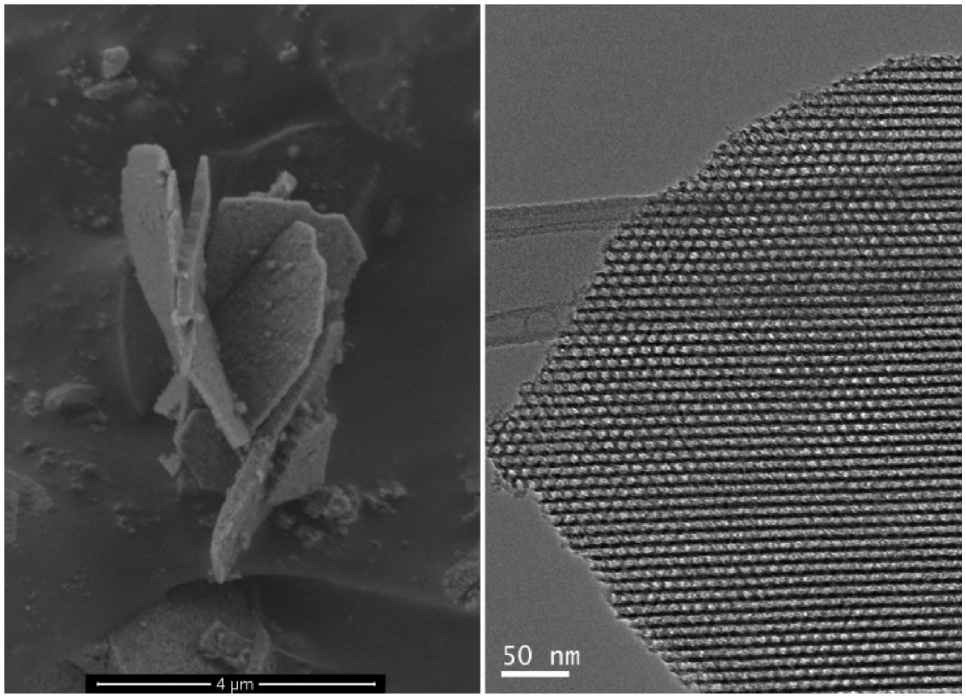


Fig. 2. Left: SEM micrograph showing the plate-like morphology of SBA-15 silica particles. Right: TEM image recorded in the direction of the axis of SBA-15 mesopore channels. For experimental details see Experimental Section.

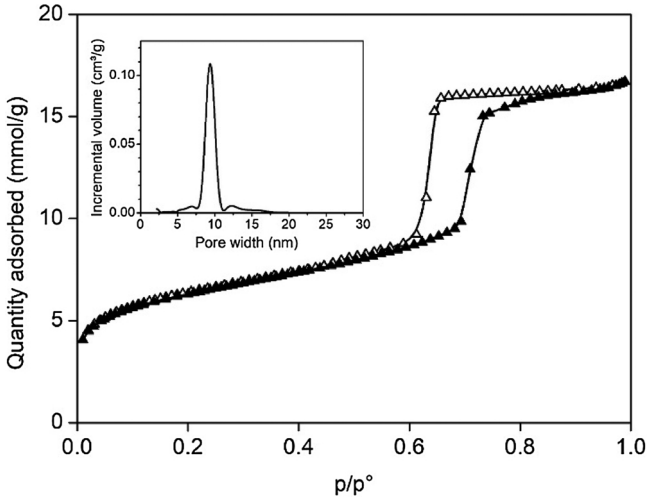


Fig. 3. Nitrogen adsorption-desorption isotherm at 77 K of silica SBA-15. Full symbols correspond to the adsorption branch and empty symbols to desorption. The inset shows the pore-size distribution calculated by NLDFT. The Nitrogen adsorption-desorption isotherm for NH_2-SBA-15 is not shown but it is estimated that the grafting of aminopropyl groups on the surface of pores will decrease the pore width by ca. 2 nm.

2.4. EH1 immobilization on magnetic microparticles

Two types of commercial (PerkenElmer LAS, Rodgau, Germany) magnetic microparticles with differentiated structural, chemical and physical properties were used for surface immobilization of the enzyme. The first one is based on a *N*-hydroxysuccinimidyl (NHS) ester modified polyvinyl alcohol (PVA) network with embedded magnetite (M) (http://www.chemagen.com/magneticbeads.html) with a diameter size range of 1–5 µm. The polymeric shell is chemically modified with the iminodiacetic acid (IDA) tridentate ligand, which is eventually used for enzyme immobilization. These beads, herein referred to as M-PVA-IDA@Ni, were activated with $NiSO_4$ prior to being used to capture His-tagged proteins. The second type of microparticles, referred to as M-Ag-NTA@Ni, is fabricated by coating a ferromagnetic core with 6% of crosslinked agarose (http://www.cube-biotech.com/s-products/magnetic-beads/his-affinity-magbeads/ni-nta-magbeads). These ready-to-use Ni-activated particles possess significantly larger dimensions (25–30 µm), and bear nitrilotetracetic acid (NTA) tetradentate ligand as chelating agent instead of IDA. The characteristics of the magnetic particles, which are commercially available, are described in https://cube-biotech.com/his-affinity-purification-guide. The particles are Ni^{2+}-activated and thus can be used for protein purification given the affinity for hexahistidine-tags. In our study we used these carriers for site-directed immobilization by capturing His-tagged proteins though Ni^{2+} which is differentially incorporated to the particles, as described above. We would like to mention that after immobilization via the hexahistidine-tags, it is possible that other groups others than His interact with the metal, or that some ion exchange may occur. The same enzyme immobilization protocol was used for both sorts of supports. Shortly, particles (50 µL; delivered as 25% particle suspensions) were retained by a magnet and the solution was decanted; after a careful removal of the supernatant the pellet was washed twice with sodium phosphate buffer (30 mM, pH 8.0) and mixed afterwards with the enzyme solution (600 µL, 1.0 mg mL^{-1}) in the same buffer. Microparticles were stirred until they were fully suspended. Upon incubation for 2 h at room temperature, the particles were once again retained by a magnet and the solution decanted and washed by repeating the procedure described above 3 times. Final enzyme-particle conjugates were suspended in 200 µL of sodium phosphate buffer (30 mM, pH 8.0) buffer and used directly for activity measurements. Protein loading was calculated by measuring the difference in protein concentration before and after the immobilization experiment. Hence, protein concentrations of $\approx$ 0.64 g L^{-1} (or $\approx$ 10.2 µg g^{-1}) for M-PVA-IDA@Ni and $\approx$ 0.52 g L^{-1} (or $\approx$ 8.3 µg g^{-1}) for M-Ag-NTA@Ni were measured.

Fig. 5 summarizes the protocol used for EH1 immobilization on magnetic microparticles. The EH1 employed was initially designed with a His-tag for purification purposes and could therefore be directly employed for immobilization.

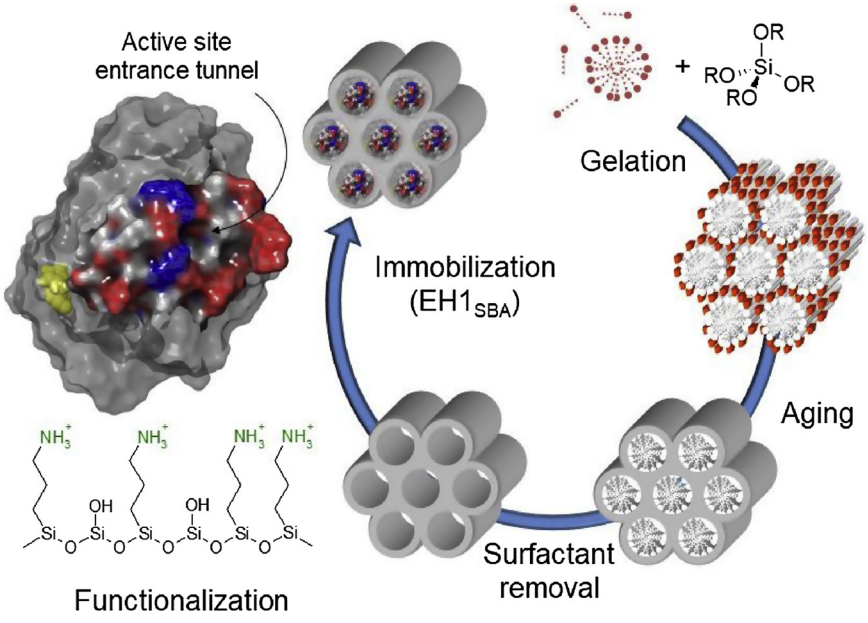


Fig. 4. Schematic representation of the protocol used for the synthesis of the NH2-SBA-15 mesostructured silica support and EH1 immobilization. EH1 structure with the cavity electrostatic surface, in red the negative charge and in blue the positive charge, created with Maestro, is represented. The access to the active site tunnel is indicated with an arrow. The position of the Hig6-tag is highlighted in yellow. Note that only one enzyme biomolecule is visible in the entrance of the pores, but each cylindrical pore channel could accommodate on average 26 enzyme molecules arranged in a row along the channel.

2.5. Ester bond hydrolysis activity assessment

Hydrolytic activity of free and immobilized preparations was assayed using a pH indicator assay at 550 nm using 96 structurally diverse esters in 384-well plates as previously described [51,52]. Briefly, reactions (total volume of 44 µL) were performed in 5 mM N-(2-hydroxyethyl)piperazine-N'-(3-propanesulfonic acid) buffer (pH 8.0) containing 1.14 mg mL^{-1} of each ester and 0.45 mM phenol red (used as pH indicator), and 2 µL of immobilized preparation (from a 10 mg solid particle per mL in 40 mM HEPES buffer pH 7.0) or free enzyme solution (1 mg mL^{-1}) were immediately added to each well using an Eppendorf Repeater M4 pipette (Eppendorf, Hamburg, Germany). The total reaction volume was 44 µL. Note that a fixed concentration of 1.14 mg mL^{-1} of each ester was used; this corresponds to an ester concentration ranging from 1.28 to 13.2 mM, depending on the ester [51], concentrations that in all cases were above the K_m values for the target enzyme [51], so that the substrate saturation was guarantee for activity tests. After incubation at 30 °C and 150 rpm in a Synergy HT Multi-Mode Microplate Reader, ester hydrolysis was measured spectrophotometrically in continuous mode at 550 nm over 24 h. One unit (U) of enzyme activity was defined as the amount of free enzyme or enzyme bound to the carrier required to transform 1 µmol of substrate in 1 min under the assay conditions using the reported extinction coefficient (εPhenol red at 550 nm = 8450 M^{-1} cm^{-1}) [51,52]. All values were corrected for non-enzymatic transformation using as controls the suspensions of the mesostructured silica and magnetic microparticles without immobilized enzyme. Note that, using those control materials, no appreciable (below detection limit; see [51]) hydrolysis was detected for any of the tested esters.

The assay used to determine the substrate specificity and enantiospecificity is pH-based, which means that the pH of the reaction solution changes as the substrates are hydrolyzed. As the immobilization strategies are sensitive towards pH, we determined the pH change during the course of the reactions to evaluate the risk of enzyme

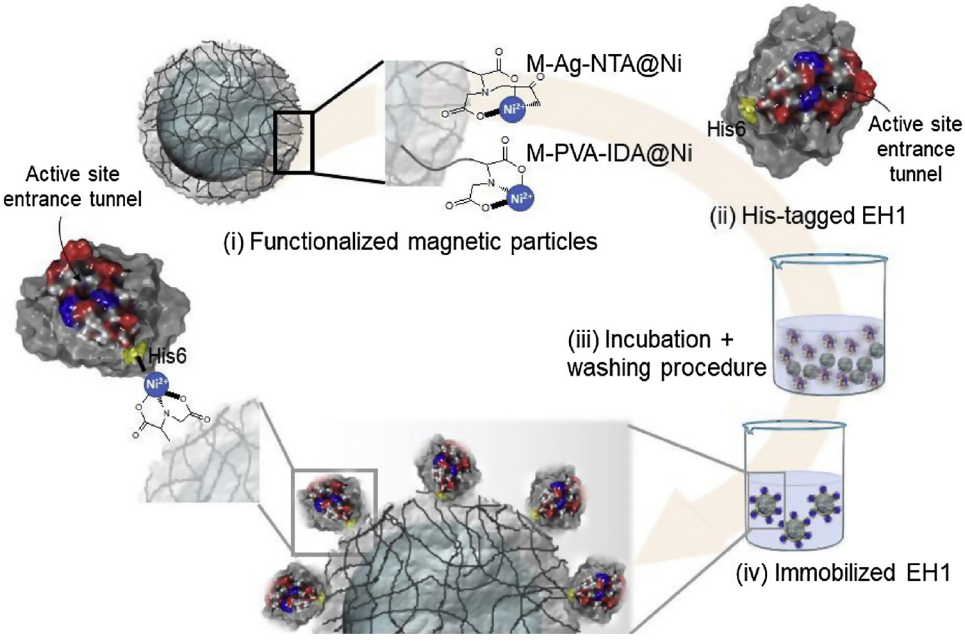


Fig. 5. Schematic representation of the protocol used for the immobilization of EH1 on two types of magnetic microparticles with different chemical and physical properties, namely, agarose-coated ferromagnetic core-shell microparticles in which EH1 was attached through a NTA linker and polyvinyl alcohol particles embedded with magnetite and grafted with an IDA linker. By using Maestro, the position of the hexahistidine-tag in EH1 is highlighted in yellow, in blue (positive charge) and red (negative charge) the residues conforming the entrance to the active site, and in grey color the rest of the protein. The access to the active site tunnel is indicated with an arrow.

leakage through pH change during the course of the reaction. Additional details on the evaluation of pH change before, during and after activity tests and their possible effect of enzyme leaching from the carriers, which was found below detection limit, are given in the Supporting Information.

2.6. Structure analysis

The crystal structure of EH1 protein was recently solved and X-ray diffraction data collection and refinement statistics are available [51]. The Maestro interface (version 10.4) was used to provide structural analysis.

3. Results and discussion

3.1. Substrate profile and enantiospecificity of EH1 immobilized in NH_2-SBA-15

EH1, which was used as model enzyme was immobilized on the amino-functionalized ordered mesoporous material NH_2-SBA-15 (see Experimental section and Fig. 4). The immobilized preparation, with enzyme loads (48 mg protein per gram of silica) in the range of previously reported studies [19,20], will be referred to as $EH1_{SBA}$. The hydrolytic activity of $EH1_{SBA}$ was evaluated relative to a set of 72 esters hydrolyzed by free EH1 (see Experimental section). The immobilization caused a significant reduction in the substrate range (Fig. 6; Table S1) as $EH1_{SBA}$ was only able to hydrolyze 17 esters, with cyclohexyl butyrate being the preferred substrate (5.3 ± 0.02 U mg^{-1}). Thus, the immobilization of the enzyme inside pores narrows the substrate scope. Most likely, this significant reduction of substrates may be a consequence of the hydrophilic surface chemistry of the functionalized mesoporous silica and of the limited access of molecules to the active site when the enzyme is immobilized inside the pores, as it has been previously observed for other enzymes [46,47]. However, the fact that $EH1_{SBA}$ efficiently hydrolyzed large molecules such as phenyl acetate and phenyl propionate, vinyl benzoate, benzyl 2-hydroxy-3-phenylpropionate, benzoic acid 4-formyl-phenylmethyl ester, 1-naphthyl acetate, 1-naphthyl butyrate, and triglycerides (Fig. 6), suggests further factors contributing to the substrate spectra of the final enzymatic preparation. Indeed, the partitioning coefficient (LogP value) and the molecular volume, which reflects electronic and steric effects and hydrophobic and hydrophilic characteristics, revealed that esters being hydrolyzed by $EH1_{SBA}$ ranged from LogP values of −0.19 to 3.85, and molecular volumes of 115.37 to 297.46 $Å^3$, respectively. However, many other esters with similar values were not accepted as substrates (Fig. 6; Table S1). Even very hydrophilic and small esters such as Υ-valerolactone (LogP: −0.27; volume: 96.74 $Å^3$) and methyl lactate (LogP: −0.72; volume: 98.57 $Å^3$) were not hydrolyzed.

Focusing on stereochemistry, EH1 was capable of hydrolyzing 14 chiral esters, that is, (*R*) and (*S*) enantiomers of menthyl acetate, methyl mandelate, ethyl 4-chloro-3-hydroxybutyrate, methyl 3-hydroxybutyrate, methyl 3-hydroxyvalerate, methyl lactate and ethyl lactate (Fig. 6; Table S1). The E_{app} for the free EH1, calculated as the ratio of specific activity (U mg^{-1}) of the preferred chiral ester over that of the non-preferred one when both esters were tested separately [53,54], ranged from 1.45 to 14.2 (Table 1). $EH1_{SBA}$ was capable of hydrolyzing only 4 out of the 14 chiral esters for which the free EH1 showed activity. Interestingly, the E_{app} for ethyl (*R*)-4-chloro-3-hydroxybutyrate over ethyl (*S*)-4-chloro-3-hydroxybutyrate (LogP: 0.33; volume: 145.95 $Å^3$) increased from 1.45 for EH1 to 419 for $EH1_{SBA}$. Slight increase in the E_{app} (2.6 vs 4.7 in the same order) was observed for ethyl (*R/S*)-lactate (LogP: -0.19; volume: 115.37 $Å^3$) (Table 1). Other chiral esters with lower or higher LogP values and molecular volumes were not accepted as substrates.

To conclude, the immobilization in NH_2-SBA-15 caused not only a significant reduction of the substrate spectrum but also a significant increase in E_{app}, particularly for ethyl (*R*)-4-chloro-3-hydroxybutyrate, a chiral compound broadly appreciated for the synthesis of biologically and pharmacologically important materials such as (*R*)-carnitine, (*R*)-4-amino-3-hydroxybutyric acid, and (*R*)-4-hydroxy-2-pyrrolidone [61]. We further evaluated whether the observed changes in specificity and E_{app} of EH1 induced by the immobilization in NH_2-SBA-15 were specific or not for this support. Therefore, we immobilized EH1 on other carriers, i.e., magnetic microparticles, where the enzyme was chemically

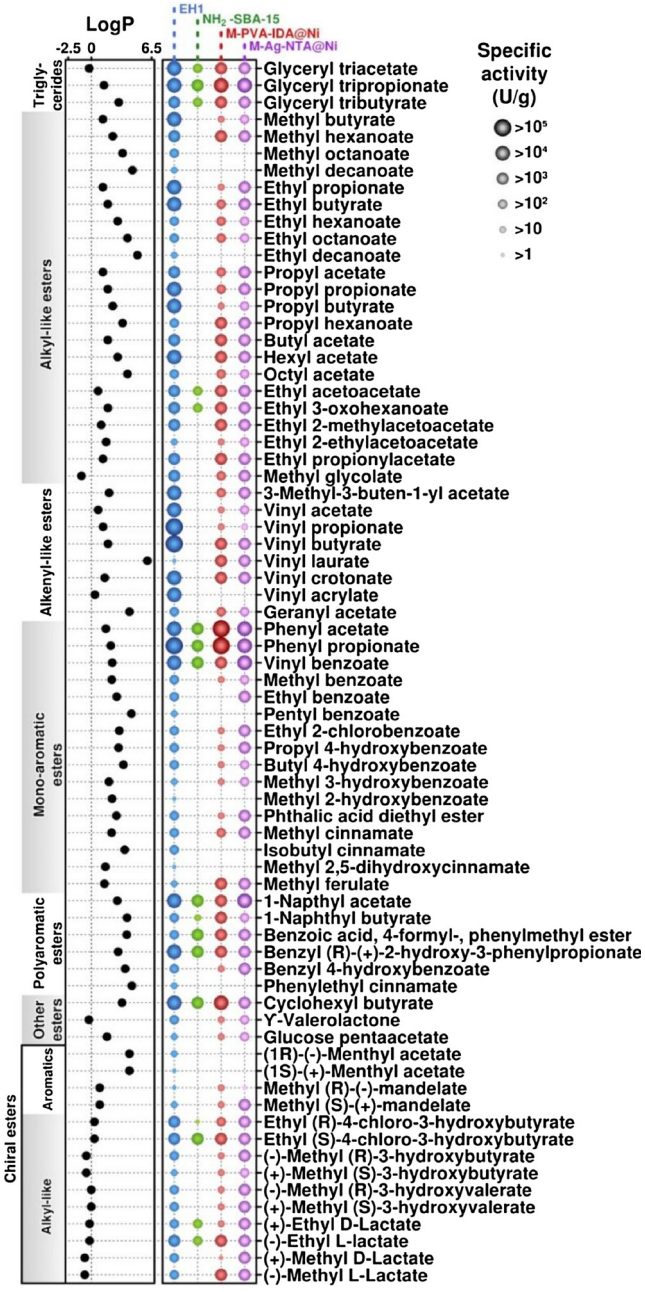


Fig. 6. Substrate range of free EH1 and its immobilized preparations. The ID code representing each variant is color coded. This Figure is created from data given in Supporting Table S1. The classes of esters tested are indicated on the left side. The Figure was created with the R language console using information about the specific activity (units g^{-1}) of the analyzed enzymes against the 72 substrates hydrolyzed by EH1 as a starting point. The activity protocol established and used to identify the esters hydrolyzed by each EH1 variant was based on a continuous pH indicator assay at pH 8.0 and 30 °C, performed in triplicate and corrected for background signal (see Experimental section). The LogP values for each ester were calculated using the ACD/ChemSketch 2015.2.5 software.

Table 1
Enantiospecificity of EH1 and immobilized preparations against a number of chiral esters, as revealed by calculations of E_{app}.

Ester[a]	Estimated E_{app}[b]			
	EH1	$EH1_{SBA}$	$EH1_{IDA}$	$EH1_{NTA}$
1**RS**	9.65 ± 0.70	n.d	n.d	n.d
2**RS**	14.2 ± 0.46	n.d	0.95 ± 0.20	151 ± 1
3**RS**	1.45 ± 0.05	419 ± 9	3.58 ± 0.03	0.64 ± 0.24
4**RS**	4.89 ± 0.36	n.d	9.09 ± 0.20	3.16 ± 0.31
5R**S**	4.41 ± 0.04	n.d	3.12 ± 0.07	1.58 ± 0.29
6R**S**	2.60 ± 0.03	4.73 ± 0.17	154 ± 9	2.02 ± 0.17
7R**S**	4.56 ± 0.36	n.d	194 ± 2	1.45 ± 0.19

[a] Compounds ID: 1RS, (*R/S*)-menthyl acetate; 2RS: methyl (*R/S*)-mandelate; 3RS: ethyl (*R/S*)-4-chloro-3-hydroxybutyrate; 4RS: methyl (*R/S*)-3-hydroxybutyrate; 5RS: methyl (*R/S*)-3-hydroxyvalerate; 6RS: ethyl (*R/S*)-lactate; 7RS: methyl (*R/S*)-lactate; the preferred chiral ester is indicated **underlined and in bold**.

[b] E_{app} was calculated per each of the pairs of enantiomers as the ratio of specific activity (U mg^{-1}) per each of the 14 chiral esters when pure stereoisomers were tested separately, as described elsewhere [53,54]. These calculations were extracted from data in Table S1. n.d.: Activity not detected under our assay conditions.

linked to the surface of the support. Other carriers such as Celite® and Sepabeads were used but the resulting preparations showed residual activity and thus were not further considered in this study.

3.2. Substrate profile and enantiospecificity of EH1 immobilized on magnetic microparticles

EH1 was immobilized on two types of magnetic microparticles (see Experimental section), namely, agarose-coated core-shell ferromagnetic particles to which EH1 was attached through a NTA linker (referred to as $EH1_{NTA}$; Fig. 4), and PVA particles with embedded magnetite and grafted with an IDA linker (referred to as $EH1_{IDA}$; Fig. 5).

$EH1_{IDA}$ preparation has the ability to hydrolyze 60 substrates, with phenyl acetate being the preferred ester (104.0 ± 0.13 U mg^{-1}) (Fig. 6; Table S1). $EH1_{NTA}$ hydrolyzed the same set of 60 substrates hydrolyzed by $EH1_{IDA}$ plus ethyl benzoate, with also phenyl acetate being the preferred ester (90.5 ± 1.8 U mg^{-1}). Compared to the free enzyme these preparations were unable to hydrolyze 11 esters, which included methyl octanoate, methyl decanoate, ethyl decanoate, pentyl benzoate, methyl 2-hydroxybutyrate, isobutyl cinnamate, phenethyl cinnamate, isobutyl cinnamate, methyl 2,5-dihydroxycinnamate, menthyl acetate, and vinyl acrylate. None of these 11 esters were either hydrolyzed by $EH1_{SBA}$. All but vinyl acrylate (LogP: 0.38; volume: 96.07 $Å^3$) could be considered as large hydrophobic esters (LogP: from 1.51 to 4.96; volume: from 136.59 to 244.4 $Å^3$). However, other large molecules with LogP values as high as 3.64 such as benzyl 4-hydroxybenzoate, and molecular volume as high as 338.38 $Å^3$ such as glucose pentaacetate, were hydrolyzed, suggesting again that factors others than surface chemistry and substrate access limitations may contribute to the substrate spectra of the final enzymatic preparations $EH1_{NTA}$ or $EH1_{IDA}$. Following on from this, it should be mentioned that both preparations showed similar substrate spectra, with a clear distinct preference for alkyl esters. Indeed, whereas $EH1_{NTA}$ showed from 3 to 19-fold higher preference for short alkyl esters such as methyl butyrate, ethyl propionate, ethyl butyrate, propyl acetate, propyl propionate, propyl butyrate, and butyl acetate, $EH1_{IDA}$ preferred (from 1.4 to 5-fold) longer esters such as methyl hexanoate, ethyl hexanoate, ethyl octanoate, hexyl acetate, and octyl acetate (Table S1).

All esters hydrolyzed by $EH1_{SBA}$ were also hydrolyzed by $EH1_{IDA}$ and $EH1_{NTA}$, whereas the immobilization on the surface of magnetic microparticles produced biocatalysts capable of converting 44 esters which $EH1_{SBA}$ did not accept. These range from small hydrophilic esters such as methyl glycolate (LogP: -1.07; volume: 81.98 $Å^3$) to large hydrophobic esters such as vinyl laurate (LogP: 6.04; volume: 252.91 $Å^3$).

With regards to stereochemistry, $EH1_{IDA}$ catalyzed the hydrolysis of all those chiral esters hydrolyzed by the free enzyme except (*R*)-(-)-menthyl acetate and (*S*)-(+)-menthyl acetate, which were not accepted as substrates (Fig. 6; Table S1). Interestingly, the E_{app} for (-)-methyl (*S*)-lactate over (+)-methyl (*R*)-lactate increased from ~4.6 for the free enzyme to ~194 for $EH1_{IDA}$ (Table 1). Similarly, the E_{app} for (-)-ethyl (*S*)-lactate over (+)-ethyl (*R*)-lactate increased from ~2.6 to ~154, in the same order. No significant changes of the E_{app} were observed for the other chiral esters, including ethyl (*R/S*)-4-chloro-3-hydroxybutyrate for which the E_{app} significantly increased after immobilization in NH_2-SBA-15. $EH1_{NTA}$ was capable of hydrolyzing the same set of chiral esters as $EH1_{IDA}$ (Fig. 4; Table S1), but a notable difference was observed. Particularly, the E_{app} for methyl (*S*)-(+)-mandelate over methyl (*R*)-(-)-mandelate increased from ~14.2 for the free enzyme to ~150 for the $EH1_{NTA}$ preparation (Table 1), substrates for which $EH1_{IDA}$ did not show any preference (E_{app} approx. 1). No significant changes of the E_{app} were observed for the other chiral esters.

In summary, immobilization on the surface of agarose-coated large core-shell ferromagnetic microparticles (25–30 μm) through a NTA linker slightly altered the substrate range (61 esters) while producing a biocatalyst more selective for short alkyl esters and for methyl (*S*)-(+)-mandelate. Moreover, immobilization on the surface of small polyvinyl alcohol magnetic microparticles (1–5 μm) through an IDA linker also slightly altered the substrate range (60 esters) while producing a more selective biocatalyst for ethyl and methyl (*S*)-lactate and slight preference for longer alkyl esters.

3.3. Flexibility constrains as determinants defining chiral preference of biocatalysts

Based on the crystal packing (PDB 5JD4) EH1 is a dimer [51], the dimensions of that are approximately 70 Å × 44 Å × 42 Å (or 7 nm × 4.4 nm × 4.2 nm). Enzymes with these dimensions have been successfully immobilized in large-pore size SBA-15 materials [38]. The mesoporous silica used to immobilize EH1 possesses pore channels which are about 200 nm long with a highly uniform diameter close to 9.3 nm (Figs. 2 and 3). The grafting of aminopropyl groups on the surface of pores makes the actual pore width available for enzyme immobilization even slightly smaller (ca. 2 nm). Therefore, it is expected that the cylindrical pore channels may accommodate only one enzyme molecule along the channel diameter. Considering the enzyme diameter (about 7 nm × 4.4 nm × 4.2 nm) and the length (200 nm) and diameter (9.3 nm for SAB-15 or ca. 7.3 nm for the amino funcionalized material) of the channels, it follows that each cylindrical pore channel could accommodate on average 26 enzyme molecules arranged in a row along the channel, in which each enzyme molecule may have restricted movement due to the similar size of enzyme structure and pores. Hence, there will be very little space for substrate and product molecules to diffuse in and out of the channels. Additional details on the enzyme load and enzyme immobilization kinetics in the NH_2-SBA-15 support are given in the Supporting Information.

To investigate which zones of the negatively charged proteins may most likely interact with the positively charged surface of the pores, a surface electrostatic analysis of the EH1 structure was performed. The analysis shows a negative zone around the binding cavity entrance (Fig. 4), and thus it is plausible that the active site is oriented to the surface of the amino-functionalized ordered mesoporous material. The hindered movement of the enzyme inside the pores and the most likely unfavorable orientation of the EH1 binding cavity may explain the restricted substrate spectrum of the $EH1_{SBA}$ preparation. However, both reasons per se do not explain the fact that the enzyme immobilized inside the 7.3 nm diameter pores of NH_2-SBA-15 materials retains the capacity to hydrolyze some very large and hydrophobic molecules such

as benzoic acid 4-formyl-phenylmethyl ester, while being unable to hydrolyze very small esters such as vinyl acetate (Fig. 6). It is plausible that factors others than access of molecules to the pores and to the active site play an additional role in determining the substrate specificity (including enantiospecificity). In this study we hypothesized that protein immobilization could affect not only the free movement of the enzyme but also impose flexibility constraints to the enzyme molecule. Note that catalytic triads in serine ester hydrolases are located in a catalytic environment where residues lining the active site are contributing to binding and correct positioning of substrates [51,52]. Altering the position, distances and angles of those residues may alter the binding capacity. This may explain the binding preference of $EH1_{SBA}$ for ethyl (*R*)-(-)-4-chloro-3-hydroxybutyrate over (*S*)-(+)-4-chloro-3-hydroxybutyrate.

To prove whether the flexibility constraints may explain the improved E_{app} for ethyl (*R*)-(-)-4-chloro-3-hydroxybutyrate, EH1 was immobilized in NH_2-SBA-15 materials with same surface chemistry (1.6 or 1.4 mmol aminopropyl groups per gram of support) but wider pores (pore diameter of 16.7 or 14.0 nm) (details on the synthesis and characterization of these supports are given in the Supporting Information). The specific activity of both preparations towards ethyl (*R*/*S*)-(-/+)-4-chloro-3-hydroxybutyrate was determined, and the E_{app} calculated. The E_{app} for EH1 immobilized in the material with a diameter of 16.7 nm was found to be 4.51 (*R* preference), and 4.57 (*R* preference) for that with 14.0 nm diameter. This value is slightly higher than that of the free enzyme (~1.5) but significantly lower than that obtained when the enzyme was immobilized in the material with pores of 9.3 nm diameter (~420) where EH1 has a tight fit. These results suggest that restricting the enzyme movement and/or promoting enzyme rigidity by non-covalent enzyme immobilization in pores with diameters close to that of the enzyme crystal packing may help promoting enantiospecificity.

In $EH1_{NTA}$ and $EH1_{IDA}$ a controlled immobilization via His-tag is used. Immobilization of proteins through the His-tag allows an oriented immobilization of all enzyme molecules, which is important to have a clue about the accessibility of the active site cavity with respect to the surface of the support. In this line, immobilization of the His-tagged enzyme onto nickel(II)-cyclam grafted mesoporous silica have been shown to allow tailored made adsorption [62] as this may avoid problems associated to an unfavourable orientation of the enzyme. An analysis of the EH1 structure revealed that orientation of EH1 both in $EH1_{NTA}$ and $EH1_{IDA}$ is favorable for the substrate access to the active site tunnel. Indeed, the *N*-terminus of the enzyme where the His-tag is located is oriented opposite to the active site cavity (Fig. 5). However, this favorable orientation, which may occur in both preparations, does not explain per se why the two immobilized preparations were not able to hydrolyze 12 ($EH1_{NTA}$) and 11 ($EH1_{IDA}$) esters which were hydrolyzed by the free enzyme. These molecules included large esters such as methyl octanoate, methyl decanoate, ethyl decanote, pentyl benzoate, methyl-2-hydroxybenzoate, isobutyl cinnamate, methyl 2,5-dihydroxycinnamate, phenylethyl cinnamate, and menthyl acetate, but also short esters such as vinyl acrylate, which a priori should cause less diffusional problems compared to larger esters. Also, it does not explain that both preparations were able to hydrolyze methyl benzoate (LogP: 2.2; volume: 263.3 $Å^3$) and vinyl benzoate (LogP: 2.25; 139.74 $Å^3$), whereas the slightly larger ethyl benzoate (LogP: 2.73; 145.38 $Å^3$) was only accepted by $EH1_{NTA}$ (Fig. 6; Table S1). Finally, it does not explain the increased preference of $EH1_{NTA}$ for methyl (*S*)-(+)-mandelate (LogP: 0.9; 153.42 $Å^3$) or the increased preference of $EH1_{IDA}$ for ethyl (*S*)-lactate (LogP: -0.72; 98.57 $Å^3$) and larger alkyl esters. We hypothesize that the stronger interaction of the protein through a NTA linker in $EH1_{NTA}$ compared to IDA linker in $EH1_{IDA}$ may slightly increase the rigidity of the protein and thus alter the entrance or positioning of a specific subset of esters, thus influencing substrate preference and specificity for specific set of esters, including chiral esters. It is also plausible that the smaller particles may have a higher density of proteins, which may affect their packing on the surface.

4. Conclusion

In this manuscript we studied the effect of immobilization of a model enzyme in an unprecedented manner. Through an extensive analysis of the substrate spectra, we examined in depth the contribution of surface chemistry, particle size, substrate accessibility to the active site tunnel, and flexibility constraints, driven by each immobilization strategy, on enzyme substrate specificity and stereochemistry. We found that flexibility constraints, which can be modulated through careful materials design (particle size, pore size and architecture) and functionalization (strength of the linkage), are among the most important factors determining the specificity and stereochemistry of immobilized enzymes. These constraints may most likely affect the enzyme structure and the active site configuration and in turn the access and positioning of substrates in the active site. This was herein proven by converting an ester hydrolase with broad substrate spectrum but non-enantiospecific, a common feature in natural esterases, into biocatalysts capable of converting specific molecules, including stereoisomers, depending on the immobilization strategy applied. Therefore, by controlling the immobilization strategy, the size and the functionalization of materials and interfaces employed, one can use substrate promiscuous enzymes to design biocatalysts for multiple selective catalytic routes and stereochemistry, and to produce a range of chiral molecules which are important building blocks for drug/fragrance discovery, chemical biology, and modern materials science. Finally, the present study suggests that a controlled manipulation of specificity of enzymes can be achieved to a higher extent than previously envisaged by size/chemistry-tunable materials, to produce more efficient catalysts. Studying mesoporous materials with a large constellation of pore sizes and materials with different density of reactive groups may be of interest in the future to obtain a clue about their impact in enzyme rigidity/flexibility and to design carriers with characteristics capable of producing a desired change in substrate specificity or stereochemistry. Investigation of immobilization effect by computational methods will help deepening into the structural constrains induced by different carriers, the quantification of the rigidity/flexibility, and its link with specificity (see additional comments in Supporting Results).

Acknowledgements

This project received funding from the European Union's Horizon 2020 research and innovation program Blue Growth: Unlockingthe potential of Seas and Oceans under grant agreement no. 634486 (project acronym INMARE). This research was also supported by the grants PCIN-2014-107 (within ERA NET IB2 grant nr. ERA-IB-14-030 - MetaCat), PCIN-2017-078 (within the ERA-MarineBiotech grant ProBone), BIO2014-54494-R, MAT2016-77496-R, BIO2017-85522-R, and CTQ2016-79138-R from the Spanish Ministry of Economy, Industry and Competitiveness. A.B. acknowledges the support from the Spanish Ministry of Economy, Industry and Competitiveness (MAT2017-88808-R grant), María de Maeztu Units of Excellence Programme (MDM-2016-0618), and the Diputación de Guipúzcoa for current funding in the frame of Gipuzkoa Fellows program. G.D. thanks the German Federal Ministry of Education and Research (BMBF, Grant No. 031A095C) for funding in the frame of the Molecular Interaction Engineering program (Biotechnologie 2020+). The authors gratefully acknowledge financial support provided by the European Regional Development Fund (ERDF). C.C. thanks the Spanish Ministry of Economy, Industry and Competitiveness for a PhD fellowship (Grant BES-2015-073829).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi: https://doi.org/10.1016/j.apcata.2018.08.003.

References

[1] D.N. Tran, Jr.K.J. Balkus, ACS Catal. 1 (2011) 956–968.
[2] A.S. Bommarius, M.F. Paye, Chem. Soc. Rev. 42 (2013) 6534–6565.
[3] G. Kde, R. Daiha, S.D. Angeli, R.V. de Oliveira, Almeida, PLoS ONE 10 (2015) e0131624.
[4] F. Jia, B. Narasimhan, S. Mallapragada, Biotechnol. Bioeng. 111 (2014) 209–222.
[5] Z. Zhou, M. Hartmann, Chem. Soc. Rev. 42 (2013) 3894–3912.
[6] E.P. Cipolatti, A. Valério, R.O. Henriques, D.E. Moritz, J.L. Ninow, D.M.G. Freire, E.A. Manoel, R. Fernandez-Lafuente, D. De Oliveira, RSC Adv. 6 (2016) 104675–104692.
[7] H.R. Luckarift, J.C. Spain, R.R. Naik, M.O. Stone, Nat. Biotechnol. 22 (2004) 211–213.
[8] S.-H. Jun, J. Lee, B.C. Kim, J.E. Lee, J. Joo, H. Park, J.H. Lee, S.-M. Lee, D. Lee, S. Kim, Y.-M. Koo, C.H. Shin, S.W. Kim, T. Hyeon, J. Kim, Chem. Mater. 24 (2012) 924–929.
[9] N. Brun, A. Babeau Garcia, H. Deleuze, M.-F. Achard, C. Sanchez, F. Durand, V. Oestreicher, R. Backov, Chem. Mater. 22 (2010) 4555–4562.
[10] M. Sevilla, T. Valdés-Solís, P. Tartaj, A.B. Fuertes, J. Colloid Interface Sci. 340 (2009) 230–236.
[11] M.I. Shukoor, F. Natalio, H.A. Therese, M.N. Tahir, V. Ksenofontov, M. Panthöfer, M. Eberhardt, P. Theato, H.C. Schröder, W.E.G. Müller, W. Tremel, Chem. Mater. 20 (2008) 3567–3573.
[12] H. Gustafsson, A. Küchler, K. Holmberg, P. Walde, J. Mater. Chem. B Mater. Biol. Med. 3 (2015) 6174–6184.
[13] P.B. Messersmith, M. Textor, Nat. Nanotechnol. 2 (2007) 138–139.
[14] Z.D. Knežević-Jugović, M.G. Žuža, S.M. Jakovetić, A.B. Stefanović, E.S. Džunuzović, K.B. Jeremić, S.M. Jovanović, Biotechnol. Prog. 32 (2016) 43–53.
[15] J. Britton, C.L. Raston, G.A. Weiss, Chem. Commun. 52 (2016) 10159–10162.
[16] E. Zare-Eelanjegh, D.K. Bora, P. Rupper, K. Schrantz, L. Thöny-Meyer, K. Maniura-Weber, M. Richter, G. Faccio, ACS Appl. Mater. Interfaces 8 (2016) 20432–20439.
[17] A. Rao, A. Bankar, A. Shinde, A.R. Kumar, S. Gosavi, S. Zinjarde, ACS Appl. Mater. Interfaces 4 (2012) 871–877.
[18] M.P. Conte, K.H. Lau, R.V. Ulijn, ACS Appl. Mater. Interfaces 9 (2017) 3266–3271.
[19] E. Magner, Chem. Soc. Rev. 42 (2013) 6213–6266.
[20] V. Gascón, I. Díaz, C. Márquez-Álvarez, R.M. Blanco, Molecules 19 (2014) 7057–7071.
[21] V. Gascón, C. Márquez-Álvarez, R.M. Blanco, Appl. Catal A: General 482 (2014) 116–126.
[22] A. Küchler, M. Yoshimoto, S. Luginbühl, F. Mavelli, P. Walde, Nat. Nanotechnol. 11 (2016) 409–420.
[23] E. Serra, A. Mayoral, Y. Sakamoto, R.M. Blanco, I. Díaz, Microporous Mesoporous Mater. 114 (2008) 201.
[24] J. Fan, W. Shui, P. Yang, X. Wang, Y. Xu, H. Wang, X. Chen, D. Zhao, Chemistry 11 (2005) 5391–5396.
[25] S. Hudson, E. Magner, J. Cooney, B.K. Hodnett, J. Phys. Chem. B 109 (2005) 19496–19506.
[26] J. He, Y. Xu, H. Ma, Q. Zhang, D.G. Evans, X. Duan, J. Colloid Interface Sci. 298 (2006) 780–786.
[27] D. Jung, C. Streb, M. Hartmann, Int. J. Mol. Sci. 11 (2010) 762–778.
[28] M. Vittorini, E. Dumitriu, G. Barletta, F. Secundo, Bioprocess Biosyst. Eng. 34 (2011) 247–251.
[29] J.M. Gómez, M.D. Romero, T.M. Fernández, E. Díez, Bioprocess Biosyst. Eng. 35 (2012) 1399–1405.
[30] M.M. Wan, W.G. Lin, L. Gao, H.C. Gu, J.H. Zhu, J. Colloid Interface Sci. 377 (2012) 497–503.
[31] H. Zhang, E. Xun, J. Wang, G. Chen, T. Cheng, Z. Wang, T. Ji, L. Wang, Int. J. Mol. Sci. 13 (2012) 5998–6008.
[32] N. Canilho, J. Jacoby, A. Pasc, C. Carteret, F. Dupire, M.J. Stébé, J.L. Blin, Colloids Surf. B Biointerfaces 112 (2013) 139–145.
[33] S. Li, Z. Wu, M. Lu, Z. Wang, Z. Li, Molecules 18 (2013) 1138–1149.
[34] P. Bhange, N. Sridevi, D.S. Bhange, A. Prabhune, V. Ramaswamy, Int. J. Biol. Macromol. 63 (2014) 218–224.
[35] G. Chandrasekar, M. Hartmann, V. Murugesan, J. Nanosci. Nanotechnol. 14 (2014) 2606–2613.
[36] A. Kumari, B. Kaur, R. Srivastava, R.S. Sangwan, Biochem. Biophys. Rep. 2 (2015) 108–114.
[37] P. Ramachandran, G.K. Narayanan, S. Gandhi, S. Sethuraman, U.M. Krishnan, Appl. Biochem. Biotechnol. 175 (2015) 2332–2346.
[38] J.M. Bolivar, V. Gascon, C. Marquez-Alvarez, R.M. Blanco, B. Nidetzky, Langmuir 33 (2017) 5065–5076.
[39] P. Corell Escuin, A. García-Bennett, J.V. Ros-Lis, A. Argüelles Foix, A. Andrés, Food Chem. 217 (2017) 360–363.
[40] X. Fan, W. Liang, Y. Li, H. Li, X. Liu, Microb. Cell Fact. 16 (2017) 149.
[41] S. Hüttner, M. Zezzi Do Valle Gomes, L. Iancu, A. Palmqvist, L. Olsson, Bioresour. Technol. 239 (2017) 57–65.
[42] A. Vinu, N. Gokulakrishnan, V.V. Balasubramanian, S. Alam, M.P. Kapoor, K. Ariga, T. Mori, Chemistry 14 (2008) 11529115–11529138.
[43] H. Gustafsson, C. Thörn, K. Holmberg, Colloids Surf. B Biointerfaces 87 (2011) 464–471.
[44] J. Lee, J. Kim, J. Kim, H. Jia, M.I. Kim, J.H. Kwak, S. Jin, A. Dohnalkova, H.G. Park, H.N. Chang, P. Wang, J.W. Grate, T. Hyeon, Small (2005) 744–753.
[45] S. Moreno-Pérez, A.H. Orrego, M. Romero-Fernández, L. Trobo-Maseda, S. Martins-DeOliveira, R. Munilla, G. Fernández-Lorente, J.M. Guisan, Methods Enzymol. 571 (2016) 55–72.
[46] D. Weber, A.J. Sederman, M.D. Mantle, J. Mitchell, L.F. Gladden, Phys. Chem. Chem. Phys. 12 (2010) 2619–2624.
[47] A. Salis, D. Meloni, S. Ligas, M.F. Casula, M. Monduzzi, V. Solinas, E. Dumitriu, Langmuir 21 (2005) 5511–5516.
[48] C. Mateo, J.M. Palomo, G. Fernandez-Lorente, J.M. Guisan, R. Fernandez-Lafuente, Enzyme Microb. Technol. 40 (2007) 1451–1463.
[49] R.C. Rodrigues, C. Ortiz, A. Berenguer-Murcia, R. Torres, R. Fernández-Lafuente, Chem. Soc. Rev. 42 (2013) 6290–6307.
[50] F. Secundo, Chem. Soc. Rev. 42 (2013) 6250–6261.
[51] M. Martínez-Martínez, C. Coscolín, G. Santiago, J. Chow, P. Stogios, R. Bargiela, C. Gertler, J. Navarro-Fernández, A. Bollinger, S. Thies, C. Méndez-García, A. Popovic, G. Brown, T.N. García Chernikova, A. García-Moyano, G.E.K. Bjerga, P. Pérez-García, T. Hai, M.V. Del Pozo, R. Stokke, I.H. Steen, H. Cui, X. Xu, B. Nocek, M. Alcaide, M. Distaso, V. Mesa, A.I. Peláez, J. Sánchez, P.C.F. Buchholz, J. Pleiss, A.F. Fernández-Guerra, F.O. Glöckner, O.V. Golyshina, M.M. Yakimov, A. Savchenko, K.-E. Jaeger, A.F. Yakunin, W.R. Streit, P.N. Golyshin, V. Guallar, M. Ferrer, ACS Chem. Biol. 13 (2018) 225–234.
[52] G. Santiago, M. Martínez-Martínez, S. Alonso, R. Bargiela, C. Coscolín, P.N. Golyshin, V. Guallar, M. Ferrer, Biochemistry 57 (2018) 2245–2255.
[53] L.E. Janes, C. Löwendahl, R.J. Kazlauskas, Chem. Eur. J. 4 (1998) 2317–2324.
[54] C. Coscolín, M. Martínez-Martínez, J. Chow, R. Bargiela, A. García-Moyano, G.E.K. Bjerga, A. Bollinger, R. Stokke, I.H. Steen, O.V. Golyshina, M.M. Yakimov, K.-E. Jaeger, A.F. Yakunin, W.R. Streit, P.N. Golyshin, M. Ferrer, Catalysts 8 (2018) 10.
[55] U.K. Laemmli, Nature 227 (1970) 680–685.
[56] M.M. Bradford, Anal. Biochem. 72 (1976) 248–254.
[57] P. Linton, V. Alfredsson, Chem. Mater. 20 (2008) 2878–2880.
[58] M. Thommes, K. Kaneko, A.V. Neimark, J.P. Olivier, F. Rodriguez-Reinoso, J. Rouquerol, K.S.W. Sing, Pure Appl. Chem. 87 (2015) 1051–1069.
[59] M.L. Ojeda, J.M. Esparza, A. Campero, S. Cordero, I. Kornhauser, F. Rojas, Phys. Chem. Chem. Phys. 5 (2003) 1859–1866.
[60] P.I. Ravikovitch, S.C.Ó. Domhnaill, A.V. Neimark, F. Schüth, K.K. Unger, Langmuir 11 (1995) 4765–4772.
[61] H. Yamamoto, A. Matsuyama, Y. Kobayashi, Biosci. Biotechnol. Biochem. 66 (2002) 481–483.
[62] D.A. Gaffney, S. O'Neill, M.C. O'Loughlin, U. Hanefeld, J.C. Cooney, E. Magner, Chem. Commun. 46 (2010) 1124–1246.

Chapter 5: Rational engineering of multiple active sites in an ester-hydrolase.

In this Chapter, we applied the knowledge recorded in Chapters 1 and 2 for tuning, by protein engineering, the properties and performance of the most substrate promiscuous esterase-hydrolase in our collection. As we have accumulated a great amount of experimental data of the topology of the ester-hydrolases binding pocket, we decided to perform an alternative engineering way to the common one consisting in random mutagenesis; rather, we used a new approach consisting in introducing a second artificial active site which can work together and separately with the native one.

The idea is to start with enzymes with proven substrate promiscuity and use this as a lead scaffold to engineer better variants in terms of expanded substrate specificity and selectivity. The EH1 ester-hydrolases was selected for the same reasons we selected it in Chapter 4, namely it is the most promiscuous ester-hydrolase but it is not enantio-selective. In this work, we have applied computational docking techniques to screen the surface of the enzyme with a typical ester substrate, glyceryl tributyrate, in order to find an alternative substrate binding site and transform it into a catalytic one by performing adequate mutations.

For this purpose, we have used the PELE (Protein Energy Landscape Exploration) software which can simulate efficiently this mapping process giving us the energy levels of each approaching of the substrate through the surface of the protein. Once we have it, we performed site-directed mutagenesis for introducing a new catalytic triad.

As a result, we succeed in designing the first ester-hydrolase with two active sites, one being the native, and one the newly introduced. As both native and artificial sites have different environments and catalytic configurations they showed different catalytic profiles. Although the expected improvement in catalytic performance and enantio-selectivity after incorporating the second site were not as expected, this article is the first one in which a second catalytic triad and oxyanion hole are successfully introduced into a native esterase already containing one. Having said that, we suggest that this first example opens a new range of possibilities for obtaining a new generation of artificial industrially-relevant biocatalysts.

Note: I am the fifth author of this article. I have participated actively in all the experimental activities such as expression, purification and characterization of the model enzyme, as said in Chapter 1. I also have participated actively in the design of mutants by computational analyses with PELE software, as I was learning the management of the PELE software in a short-stay in the Barcelona Supercomputing Center.

Biochemistry

Article

Cite This: Biochemistry 2018, 57, 2245−2255

pubs.acs.org/biochemistry

Rational Engineering of Multiple Active Sites in an Ester Hydrolase

Gerard Santiago,[†,○] Monica Martínez-Martínez,[‡,○] Sandra Alonso,[‡] Rafael Bargiela,[‡,⊥] Cristina Coscolín,[‡] Peter N. Golyshin,[§,∥] Víctor Guallar,[*,†,#] and Manuel Ferrer[*,‡]

[†]Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain

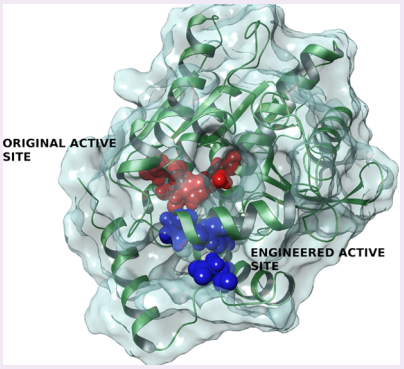
[‡]Institute of Catalysis, Consejo Superior de Investigaciones Científicas, 28049 Madrid, Spain

[⊥]School of Chemistry, [§]School of Biological Sciences, and [∥]Centre for Environmental Biotechnology, Bangor University, LL57 2UW Bangor, United Kingdom

[#]Institucio Catalana de Recerca i Estudis Avancats (ICREA), 08010 Barcelona, Spain

Supporting Information

ABSTRACT: Effects of altering the properties of an active site in an enzymatic homogeneous catalyst have been extensively reported. However, the possibility of increasing the number of such sites, as commonly done in heterogeneous catalytic materials, remains unexplored, particularly because those have to accommodate appropriate residues in specific configurations. This possibility was investigated by using a serine ester hydrolase as the target enzyme. By using the Protein Energy Landscape Exploration software, which maps ligand diffusion and binding, we found a potential binding pocket capable of holding an extra catalytic triad and oxyanion hole contacts. By introducing two mutations, this binding pocket became a catalytic site. Its substrate specificity, substrate preference, and catalytic activity were different from those of the native site of the wild type ester hydrolase and other hydrolases, due to the differences in the active site architecture. Converting the binding pocket into an extra catalytic active site was proven to be a successful approach to create a serine ester hydrolase with two functional reactive groups. Our results illustrate the accuracy and predictive nature of modern modeling techniques, opening novel catalytic opportunities coming from the presence of different catalytic environments in single enzymes.



A number of studies have shown the utility of placing multiple functional groups randomly or uniformly on the surface of solid heterogeneous catalysts to enhance activity and selectivity.[1] Functional groups are positioned in those catalysts at varying distances and spatial arrangements. However, these advances have not been as well developed for enzymatic catalysts. Thus, with few exceptions,[2−4] all enzymes contain one active site. Although introducing single active sites into noncatalytic proteins has been successfully achieved,[5−16] no evidence has been gathered that demonstrates that introducing a second or more active sites into an enzyme is feasible. If one could use molecular modeling to discover extra binding pockets and convert them into catalytic sites, then one would facilitate designing enzymes with multiple reactive groups. This would promote the competitiveness and catalytic opportunities of enzymes coming from different catalytic environments and be a considerable step forward in the field of *de novo* enzyme design.

We tested this hypothesis by using a serine ester hydrolase from the structural superfamily of α/β-hydrolases as a model enzyme. The activity of serine ester hydrolases relies mainly on a catalytic triad usually formed by Ser, Asp/Glu, and His.[17] The initial attack of an ester is mediated by the Ser residue that acts as nucleophile through which a covalent intermediate is formed in collaboration with two other residues (Asp and His) that help activate the nucleophile by forming a charge-relay network. At basic pH the His residues are δ-protonated, and the Asp residues are deprotonated, resulting in the formation of a His-Ser and His-Asp hydrogen-bonding network. The stabilization of the intermediate is increased by so-called oxyanion hole contacts with nitrogen atoms of adjacent Gly residues.[17]

This enzyme class was selected for a number of reasons. First, it is widely distributed in the environment, it has important physiological functions, it includes hydrolases which are among the most important industrial biocatalysts, and extensive biochemical knowledge has been accumulated.[18] Second, a number of studies have disclosed the utility of computational tools to introduce active sites supporting ester hydrolysis in a number of noncatalytic proteins. A first attempt used RosettaMatch to search for catalytically inert theozyme scaffold proteins that could accommodate a model catalytic triad (Cys-His-Glu/Asp) and oxyanion holes reassembling cysteine-like geometries.[11] Authors identified four protein scaffolds with minimalistic catalytic schema consisting of a Cys nucleophile, a nearby His and a backbone NH group, which were capable of hydrolyzing the model esters coumarin-(2-phenyl)-propanoate and *p*-nitrophenyl-(2-phenyl)-propa-

Received: March 6, 2018
Published: March 30, 2018

DOI: 10.1021/acs.biochem.8b00274
Biochemistry 2018, 57, 2245−2255

noate (pN2PP) (k_2/K_S up to 309 M^{-1} s^{-1}). More recently, Moroz and colleagues took advantage of existing enzymatic folds to explore whether calmodulin (CaM) could assume esterase activity against pN2PP.[16] Docking was used to determine whether *p*N2PP accommodates into a hydrophobic cavity of CaM. Residues facing the substrate were then identified and further mutated to His, a residue that, per se, has been shown to catalyze hydrolysis of *p*N2PP, albeit with a modest activity. A mutant with a single Met144His substitution was capable of hydrolyzing *p*N2PP (k_{cat}/K_M up to 4800 M^{-1} min^{-1}). In another approach, the hydrolysis of *p*-nitrophenyl acetate (*p*NPA) was achieved by incorporating mutations favoring metal binding, with the metal being the catalytic component. By using this approach, Zn^{2+} and/or Hg^{2+} ions were introduced into the Rab-4-binding domain,[10] the TRI family of peptides,[12] and short amyloid-forming peptides.[14] The resulting metallo-proteins were capable of *p*NPA hydrolysis (k_{cat}/K_M up to 630, 23.3, and 18 M^{-1} s^{-1}). The question of where to put the functional group has been found to be critical for catalytic performance. For example, it has been demonstrated that the His orientation modulates local metal orientation, which in turn has functional consequences.[19,20]

It remains to be established whether the computational approaches described above can be implemented to introduce multiple active sites supporting ester hydrolysis, not only in noncatalytic proteins but also in ester hydrolases already containing a native functional group. The novelty of this study relies on using the Protein Energy Landscape Exploration (PELE) software. PELE offers one of the best modeling alternatives to map protein−ligand dynamics and induced fit.[21−25] It allows for a complete protein surface exploration, locating binding site pockets in only a few hours of a moderate computing cluster (~32 computing cores).[26,27] Thus, these technological developments are ideally suited to locate, without previous knowledge, potential binding sites that could be converted into active sites. Finding these pockets typically requires significant enzyme reorganization, both at the level of conformational sampling and induced fit.[27] Thus, similar analysis cannot be performed with simple docking techniques. Effective alternative techniques are mostly limited to molecular dynamic simulations,[28] although at a significantly higher computational cost.

The present study adds important insights and empirical and computational data proving, for first time to the best of our knowledge, that introducing extra catalytic reactive groups into a serine ester hydrolase is plausible. We would like to highlight that our approach is based on locating, by PELE software, extra binding pockets and converting them into catalytic active sites. We were interested not only in proving that introducing extra functional groups into a catalytic serine ester hydrolase is plausible, but also in examining the catalytic potential of the resulting variant compared to the wild type enzyme and other native ester hydrolases. In this line, we have unambiguously confirmed that the extra catalytic site is not only competitive with those of other native ester hydrolases, but also could confer catalytic changes when introduced in the wild type ester hydrolase already containing a native active site.

EXPERIMENTAL AND COMPUTATIONAL METHODS

Chemicals and Oligonucleotides. All chemicals used for enzymatic tests were of the purest grade available and were purchased from Sigma Chemical Co. (St Louis, MO, USA), Alfa Aesar (Karlsruhe, Germany) or Santa Cruz Biotechnology (Heidelberg, Germany). The oligonucleotides used for DNA amplification were synthesized by Sigma Genosys Ltd. (Pampisford, Cambs, UK).

$EH1_A$ Protein Source and Crystal Structure. The isolation of the enzyme $EH1_A$ used in the present study was reported previously.[29] The enzyme is available in the expression vector pET46 Ek/LIC plasmid and *Escherichia coli* BL21 as a host,[29,30] which was the source of the enzyme for the present study. The crystal structure of $EH1_A$ protein was recently solved (PDB code: 5JD4), and X-ray diffraction data collection and refinement statistics are available.[30]

Site Directed Mutagenesis. Mutagenic PCR was developed using the Quick Change Lightning Multi Site-Directed Mutagenesis kit (Agilent Technologies, Cheadle, UK), following manufacturer instructions. Briefly, 50 ng of pET46 Ek/LIC plasmid containing wild type $EH1_A$ DNA insert was mixed with a master mix containing 2.5 μL of 10× Multi reaction buffer, 0.5 μL of Quick solution, 1 μL of dNTP mix, 1 μL of Multi enzyme blend, and 100 ng of each primer. Distilled water was added to a final volume of 25 μL. PCR conditions were as follows: 2 min at 95 °C, 30 cycles of 20 s at 95 °C, 30 s at 55 °C, 195 s at 65 °C, and one cycle of 5 min at 65 °C. The resulting variant plasmids were transferred into *E. coli* BL21 and selected on the Luria−Bertani (LB) agar supplemented with 50 μg mL^{-1} ampicillin. To obtain the variant $EH1_B$, the following primers were used: Ser161Ala Fwd (GTG GGC GGC GAT GCG GCG GGC GGC G), Glu25Asp Fwd (CGG CCC CGG CTG GAT ACC CTG CCG CAT GC), and Leu214His Fwd (TTC CTC AGC AAG GCG CAC ATG GAC TGG TTC TGG G). To obtain the variant $EH1_{AB}$, only the primers Glu25Asp Fwd and Ile214His Fwd were used.

To obtain variants containing Ser211Ala, Ser161Ala, Asp25Gln, Asp256Gln, His214Phe, and His286Phe mutations, individually or in combination, the pET46 Ek/LIC plasmid containing $EH1_A$, $EH1_B$, or $EH1_{AB}$ DNA inserts and the primers Ser211Ala Fwd (GCC GAA GGC TAC TTC CTC GCC AAG GCG CAC ATG GAC TGG), Ser161Ala Fwd (GTG GGC GGC GAT GCG GCG GGC GGC G), Asp25Gln Fwd (CGG CCC CGG CTG CAG ACC CTG CCG CAT GC), Asp256Gln Fwd (ACC GCC GGC TAC CAA CCG CTG CGC GAC G), His214Phe Fwd (TTC CTC AGC AAG GCG TTC ATG GAC TGG TTC TGG G), and His286Phe Fwd (T CCC GGC ACC ATC TTC GGC TTC TTC TCG) were used. Mutagenic PCR conditions were as above.

Gene Expression and Protein Purification. Protein expression and purification of wild type and mutants were performed as previously described with slight modifications.[29,30] Briefly, selected *E. coli* clones that expressed each protein, His-tagged at the N-terminus, were grown at 37 °C on solid LB agar medium supplemented with 50 μg mL^{-1} ampicillin, and one colony was picked and used to inoculate 10 mL of LB broth plus antibiotic in a 0.25-l flask. The cultures were then incubated at 37 °C and 200 rpm overnight. Afterward, 10 mL of this culture was used to inoculate 0.5 L of LB medium, which was then incubated to an $OD_{600\ nm}$ to approximately 0.7 (ranging from 0.55 to 0.75) at 37 °C. Protein expression was induced by adding isopropyl β-D-1-thiogalactopyranoside to a final concentration of approximately 1 mM, followed by incubation for 16 h at 16 °C. The cells were harvested by centrifugation at 5000*g* for 15 min to yield a pellet of 2−3 g L^{-1} pellet (wet weight). The wet cell pellet was frozen at −86 °C overnight, thawed, and resuspended in 15 mL of 40

DOI: 10.1021/acs.biochem.8b00274
Biochemistry 2018, 57, 2245−2255

mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.0. Lysonase Bioprocessing reagent (Novagen, Darmstadt, Germany) was added (4 μL g^{-1} wet cells) and incubated for 60 min on ice with rotating mixing. The cell suspension was sonicated using a pin Sonicator 3000 (Misonix, New Highway Farmingdale, NY, USA) for a total time of 5 min (10 W) on ice and centrifuged at 15000*g* for 15 min at 4 °C, and the supernatant was retained.

The His-tagged proteins, native and engineered variants, were purified at 4 °C after binding to a Ni-NTA His-Bind resin (Sigma Chemical Co. (St. Louis, MO, USA)), followed by ultrafiltration through low-adsorption hydrophilic 10000 nominal molecular weight limit cutoff membranes (regenerated cellulose, Amicon) to concentrate the protein solution. An extensive dialysis of protein solutions against 40 mM HEPES buffer (pH 7.0) was then performed using Pur-A-LyzerTM Maxi 1200 dialysis kit ((Sigma Chemical Co. (St. Louis, MO, USA)), as follows. Five milliliters concentrated protein solution was dialyzed against 2 L buffer during 1 h at room temperature, after which the buffer was changed by other 2 L buffer and maintained 1 h more. Then, the buffer was changed, and the dialysis was kept overnight at 4 °C. The dialyzed protein solution was recovered and concentrated as before. The concentrated protein solution (10 mg mL^{-1}) was then subjected to size-exclusion chromatography by a fast protein liquid chromatography (FPLC) equipment (LCC-500CI, Amersham Bioscience, Barcelona, Spain). The protein sample was loaded onto the FPLC coupled with a Superdex 75 size exclusion column pre-equilibrated with buffer HEPES buffer (pH 7.0). The proteins were eluted with the same buffer at a flow rate of 1 mL min^{-1}. Fractions with hydrolytic activity were pooled, concentrated, and dialyzed against HEPES buffer (pH 7.0), as before. Throughout the purification protocol, the fractions were analyzed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) on 12% gels, in a Mini PROTEAN electrophoresis system (Bio-Rad),[31] in which the proteins were stained with Coomassie brilliant blue (Protoblue Safe, National Diagnostics, GA, USA), and for hydrolytic activity using *p*-nitrophenyl propionate (*p*NPP) as described below. The protein concentration was determined according to Bradford with bovine serum albumin as the standard.[32]

Purity was assessed as >99% by SDS-PAGE, matrix-assisted laser desorption/ionization-time-of-flight/time-of-flight (MALDI-TOF/TOF) and other complementary techniques (see Supporting Results).

Ester Bond Hydrolysis Activity Assessment: Substrate Profiling Tests with 96 Esters. Hydrolytic activity was assayed at 550 nm using structurally diverse esters in 384-well plates as previously described.[30,33] All chemicals used were of the purest grade available. Briefly, before an assay, a concentrated ester stock solution was prepared in a 96-well plate by dissolving each of the 96 esters at a concentration of 25 mg mL^{-1} in acetonitrile or dimethyl sulfoxide (DMSO), depending on its solubility. Stock solutions were prepared immediately prior to use and maintained in a 96-deep-well plate at 4 °C.

The assays were conducted according to the following steps. First, a 384-well plate (Molecular Devices, LLC, CA, USA) was filled with 20 μL of 5 mM *N*-(2-hydroxyethyl)piperazine-*N*′-(3-propanesulfonic acid (EPPS) buffer, pH 8.0, using a QFill3 microplate filler (Molecular Devices, LLC, CA, USA). Second, 2 μL of each ester stock solution was added to each well using a PRIMADIAG liquid-handling robot (EYOWN TECHNOLOGIES SL, Madrid, Spain). Each of the 96 esters was dispensed in four replicates in each 384-plate. After the esters were added, the 384-well plate was filled with 20 μL of 5 mM EPPS buffer, pH 8.0, containing 0.912 mM Phenol Red (used as a pH indicator) using a QFill3 microplate filler. The final ester concentration in each well was 1.14 mg mL^{-1}, and the final concentration of Phenol Red was 0.45 mM. A total of 2 μL of protein solution (from a 1 (for EH1$_A$/EH1$_{AB}$) or 8 (for EH1$_B$) mg mL^{-1} stock solution in 40 mM HEPES buffer pH 7.0) was immediately added to each well using an Eppendorf Repeater M4 pipet (Eppendorf, Hamburg, Germany) or a PRIMADIAG liquid-handling robot. Accordingly, the total reaction volume was 44 μL, with 4.5% (v/v) acetonitrile or DMSO in the reaction mixture. After incubation at 30 °C and 150 rpm in a Synergy HT Multi-Mode Microplate reader, ester hydrolysis was measured spectrophotometrically in continuous mode at 550 nm for a total time of 24 h. One unit (U) of enzyme activity was defined as the number of enzyme required to transform 1 μmol of substrate in 1 min under the assay conditions using the reported extinction coefficient ($\varepsilon_{\text{Phenol red}}$ at 550 nm = 8450 M^{-1} cm^{-1}).[29] All values were corrected for nonenzymatic transformation.

Kinetic Measurements. For determination of kinetic parameters, these were calculated by simple Michaelis−Menten kinetics.

Kinetics experiments for *p*NPP were initiated by the addition of a stock solution (100 mM in acetonitrile) of *p*NPP to 195 μL of HEPES buffer (pH 7.0) containing EH1$_A$, EH1$_B$, or EH1$_{AB}$ in 96-well microtiter plates. Kinetics experiments for phenyl propionate were initiated by the addition of a stock solution (100 mM in acetonitrile) of phenyl propionate to 40 μL of 5 mM EPPS buffer pH 8.0 with 0.45 mM Phenol Red (used as a pH indicator) containing EH1$_A$, EH1$_B$, or EH1$_{AB}$ in 96-well microtiter plates. For K_m determinations, the amount of protein was 0.1, 16, and 1.4 μg for EH1$_A$, EH1$_B$, and EH1$_{AB}$, respectively, or 100 μg for other mutants. For k_{cat} determinations, the amount of *p*NPP and phenyl propionate was 1 and 6 mM, respectively.

In all cases, reactions were followed at 410 nm (for *p*NPP hydrolysis) or 550 nm (for phenyl propionate hydrolysis) by UV−vis spectrophotometry in a Synergy HT Multi-Mode Microplate Reader. Initial rates were determined from linear fits of the absorbance versus time (<10% conversion) corrected for the rate of uncatalyzed hydrolysis. In all cases, one unit (U) of enzyme activity was defined as the number of enzyme required to transform 1 μmol substrate in 1 min under the assay conditions using the reported extinction coefficient ($\varepsilon_{\text{pNPP}}$ at 410 nm = 15200 M^{-1} cm^{-1}; $\varepsilon_{\text{Phenol red}}$ at 550 nm = 8450 M^{-1} cm^{-1}). In all cases, reactions (performed in triplicate) were maintained at 30 °C.

pH Optima Determination. Britton and Robinson buffer (50 mM; pH 4.0−9.5) and *p*NPP (1 mM) were used for determining optimal pH of the enzyme variants. A total of 2 μL of a stock solution (100 mM in acetonitrile) of *p*NPP was added to 195 μL of HEPES buffer (pH 7.0) containing EH1$_A$ (0.1 μg), EH1$_B$ (16 μg), or EH1$_{AB}$ (1.4 μg) in 96-well microtiter plates. The *p*NPP hydrolysis was monitored as above but at 346 nm ($\varepsilon_{\text{pNPP}}$ at 346 nm = 4800 M^{-1} cm^{-1} regardless of solution pH), in triplicates.

Protein Energy Landscape Exploration (PELE) Simulations. The initial structure was taken from the coordinates of the EH1$_A$ crystal structure (PDB code: 5JD4).[30] The protonation state of titratable residues was estimated with the

DOI: 10.1021/acs.biochem.8b00274
Biochemistry 2018, 57, 2245−2255

Protein Preparation Wizard (PROPKA)[34] and the H++ server (http://biophysics.cs.vt.edu/H++) followed by visible inspection. At pH 8 (the pH at which the activity assays were performed), the catalytic triad histidine residues were δ-protonated, and the catalytic triad aspartic acid residues were deprotonated, resulting in the formation of a histidine-serine and histidine-aspartic hydrogen-bonding network. The glyceryl tripropionate structure was fully optimized with Jaguar[27] in an implicit solvent, and the electrostatic potential charges were computed with the density functional M06 at the 6-31G* level of theory; ligand parameters were extracted from these for the classic simulations.

We used PELE software to sample the binding mode of glyceryl tripropionate with $EH1_A$.[21] PELE is a Monte Carlo algorithm composed of a sequence of perturbation, relaxation, and Metropolis acceptance tests. In the first step, the ligand is subjected to random rotations and translations, while the protein is perturbed based on the anisotropic network model (ANM).[27] The maximum allowed translation for the ligand perturbation was 1.5 Å, and the maximum rotation was 20°. During the protein perturbation, all atoms were displaced by a maximum of 0.5 Å by moving the α-carbons following a random linear combination of the six lowest eigenvectors obtained in the ANM model. The relaxation step included the repositioning of all amino acid side chains within 6 Å of the ligand and the five side chains with the highest energy increase along the previous ANM step. The relaxation stage ended with a truncated Newton minimization using the OPLS all-atom force field and an implicit surface-generalized Born continuum solvent.[26] The new proposed minima were then accepted or rejected based on a Metropolis test. The substrate binding plots contained all accepted conformations for three 12-h simulations using 200 processors.

Molecular Dynamics. 250 ns of molecular dynamics (MD) simulation with DESMOND[35] were performed to ensure the enzymatic stability. After appropriate preparation of the system, as explained before, an orthorhombic water box with a minimum distance of 10 Å was introduced. The systems were then neutralized and 150 mM NaCl added. Equilibration using the default protocol was performed followed by 20 ns NPT simulation at 300 K and 1 atm with the OPLS-2005 force field. The temperature was regulated with the Nose−Hoover chain thermostat, while the pressure was controlled by the Martyna−artynae was cbarostat with isotropic coupling and a relaxation time of 2.0 ps.

Peptide Mass Fingerprinting by Matrix-Assisted Laser Desorption/Ionization-Time-of-Flight/Time-Of-Flight (MALDI-TOF/TOF). Before MALDI-TOF/TOF analysis in-solution, protein digestion was performed.[36] Briefly, 20 μg of protein samples (5 mg mL^{-1} in 40 mM HEPES buffer pH 7.0) were diluted and denatured in 20 μL of 7 M urea/2 M thiourea/100 mM triethylammonium bicarbonate (TEAB), pH 7.5, reduced with 2 μL of 50 mM tris(2-carboxyethyl) phosphine (TCEP, AB SCIEX), pH 8.0, at 37 °C for 60 min, followed by addition of 2 μL of 200 mM cysteine-blocking reagent (methylmethanethiosulfonate (MMTS); Pierce, Appleton, WI, USA) for 10 min at room temperature. Samples were diluted up to 120 μL to reduce guanidine concentration with 50 mM TEAB. Digestions were performed using sequence grade-modified trypsin (Promega, Alcobendas, Spain) to each sample in a ratio 1/20 (w/w), which were then incubated at 37 °C overnight on a shaker. Sample digestions were evaporated to dryness and were cleaned-up/desalted using Stage-Tips with Empore 3 M C18 disks (Sigma Chemical Co.; St. Louis, MO, USA). The tryptic eluted peptides were dried by speed-vacuum centrifugation and resuspended in 4 μL of MALDI solution (30% acetonitrile/15% isopropanol/0.5% trifluoroacetic acid). A 0.8 μL aliquot of each peptide mixture was deposited onto a 384-well OptiTOF Plate (SCIEX, Foster City, CA) and allowed to dry at room temperature. A 0.8 μL aliquot of matrix solution (3 mg mL^{-1} α-cyano-4-hydroxycinnamic acid in MALDI solution) was then deposited onto dried digest and allowed to dry at room temperature.

For MALDI-TOF/TOF analysis, samples were automatically acquired in an ABi 4800 MALDI TOF/TOF mass spectrometer (SCIEX, Foster City, CA) in positive ion reflector mode (the ion acceleration voltage was 25 kV to MS acquisition and 2 kV to MSMS), and the obtained spectra were stored into the ABi 4000 Series Explorer Spot Set Manager. PMF and MSMS fragment ion spectra were smoothed and corrected to zero baseline using routines embedded in ABi 4000 Series Explorer Software v3.6. Each PMF spectrum was internally calibrated with the mass signals of trypsin autolysis ions to reach a typical mass measurement accuracy of <25 ppm. Known trypsin and keratin mass signals, as well as potential sodium and potassium adducts (+21 Da and +39 Da) were removed from the peak list. To submit the combined PMF and MS/MS data to MASCOT software v.2.6.0 (Matrix Science, London, UK), GPS Explorer v4.9 was used, searching in a custom protein database with the sequences encoding $EH1_A$, $EH1_B$ and $EH1_{AB}$. The following search parameters were used: enzyme, trypsin; allowed missed cleavages, 1; methylthiolation cysteine as fixed modification by the treatment with MMTS; variable modifications, oxidation of methionine; mass tolerance for precursors was set to ±50 ppm and for MS/MS fragment ions to ±0.3 Da. The confidence interval for protein identification was set to ≥95% ($p < 0.05$) and only peptides with an individual ion score above the identity threshold were considered correctly identified.

Estimation of Molecular Mass by MALDI-TOF/TOF. Protein samples were diluted at 1:1 ratio (v/v) with matrix solution (50% saturated sinapinic acid in 70% aqueous acetonitrile and 0.1% trifluoroacetic acid). A 1.0 μL aliquot of this mixture was manually deposited onto a 386-well OptiTOF Plate (ABSciex, Framingham, MA, USA) and allowed to dry at room temperature. For MALDI-TOF/TOF analysis, samples were automatically acquired in an ABi 4800 MALDI TOF/TOF mass spectrometer (SCIEX, Foster City, CA) in positive ion linear mode (the ion acceleration voltage was 25 kV for MS acquisition). The detection mass range was set between 1500 and 80000 m/z.

RESULTS AND DISCUSSION

Model Ester Hydrolase for This Study. We chose a serine ester hydrolase, herein referred to as $EH1_A$, with a typical α/β hydrolase fold as a model. In a recent study, it was identified as the most promiscuous ester hydrolase among a total of 147 when tested with a set of 96 chemically and structurally different esters (Supporting Table S1).[30] $EH1_A$ was isolated from the metagenomic DNA of microbial communities inhabiting a karstic lake,[29] and its structure was recently solved (PDB code 5JD4).[30] The active site is formed by a catalytic triad formed by Ser161, Asp256, and His286 (Figure 1) and an oxyanion hole formed by Gly88, Gly89, and Gly90.[30] This active site supports the hydrolysis of a broad range of 72 esters

DOI: 10.1021/acs.biochem.8b00274
Biochemistry 2018, 57, 2245−2255

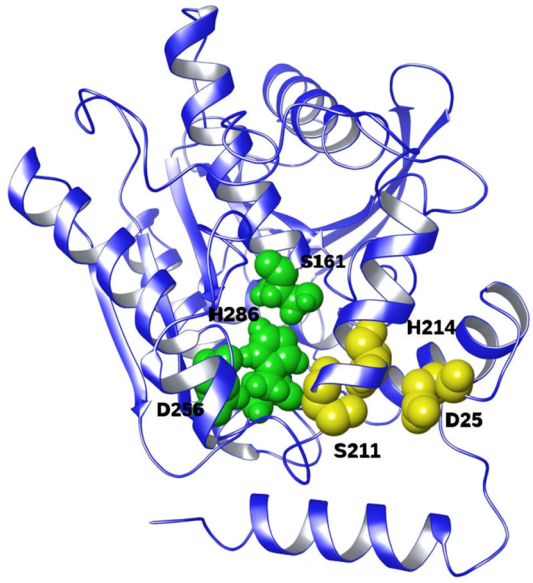


Figure 1. Relative position of the two active sites, with their catalytic triads. Original $EH1_A$ catalytic triad and oxyanion hole are shown in green and the designed $EH1_B$ catalytic triad and oxyanion hole in yellow.

(Figure 2, inset), with vinyl butyrate and phenyl propionate serving as best substrates (Figure 2).

PELE Simulations for Locating an Extra Binding Pocket. Using the PELE software, which allows mapping ligand diffusion and binding,[21−25] we performed an exploration of $EH1_A$ with glyceryl tripropionate, a ligand with high activity in multiple ester hydrolases, including $EH1_A$ (~62000 ± 7400 U g^{-1}; Figure 2), and appropriately sized for binding into defined cavities. The rationale was to identify (alternative) potential binding sites to accommodate a new active site (Figure 3). PELE simulations revealed a potential second binding site located ~13 Å from the native catalytic position at Ser161 (Figure 4a). This second site (Figure 4b) already contains a serine residue (Ser211). Thus, we computationally designed additional mutations, adding Asp and His residues to build a proper catalytic triad, taking special care of distances between the residues and substrate accommodation. PELE results for the Glu25Asp and Leu214His double mutant revealed good enzyme−substrate interaction energies (Supporting Figure S1) and a suitable catalytic position for the glyceryl tripropionate substrate. Figure 4 summarizes the catalytic triad environment of the newly introduced active site compared to the wild type. In addition, our results show that residues Gly207, Tyr208 and Phe209 act as a potential oxyanion hole, a key element in ester hydrolase catalysis (Figure 4b). Moreover, extensive MD simulations indicate proper stabilization of the double mutant (Supporting Figure S2).

Converting the Binding Pocket into a Functional Catalytic Site by Site-Directed Mutagenesis. By using site-directed mutagenesis and with the above considerations, an enzyme variant with the new presumptive active site and an inactive original site was designed to determine whether the extra active site was, per se, functional. We introduced Glu25Asp, Leu214His, and Ser161Ala substitutions, so that this variant, named $EH1_B$, would presumably employ a new catalytic triad (Ser211, Asp25, and His214) with Ser211 as the nucleophile and a new oxyanion hole (Gly207, Tyr208, and Phe209). The corresponding gene was cloned, and the His_6-tagged protein was expressed and purified by Ni-NTA affinity and size-exclusion chromatography (Supporting Figure S3 and Table S2). A number of control experiments were performed to ensure the purity and identity of the $EH1_B$ protein (Supporting Figures S3−S8). These analyses were also performed with purified $EH1_A$ and other variants (see below). With the sensitivity allowing MALDI-TOF/TOF and other complementary techniques (see Supporting Results), we can conclude that there is no contamination and that the purity of the $EH1_B$ (as well as $EH1_A$ and further variant $EH1_{AB}$) was higher than 99%. The introduced mutations were confirmed in all cases by sequencing of the corresponding gene and by peptide mass fingerprinting by MALDI-TOF/TOF of purified proteins (for details see Supporting Results, and Figures S5 and S6).

Once the purity and nature of $EH1_B$ was confirmed, its hydrolytic activity was evaluated against the set of 96 esters as for the wild-type ester hydrolase (see Experimental Section). We found that $EH1_B$ with the new active site showed not only measurable activity but also catalyzed reactions with 24 different substrates (Figure 2), including cyclohexyl butyrate and the large aromatic ester benzoic acid, 4-formyl-, phenylmethyl ester, which are rarely hydrolyzed by many native ester hydrolases (Figure 5; esters ID nos. 59 and 67).[30] Note that in $EH1_B$ the original active site (Ser161) was silenced by mutating it by Ala, the mutation per se found to completely abolish the activity of the original active site (Supporting Figure S9D). Thus, the capacity of $EH1_B$ to convert these 24 esters could be unambiguously assigned to the new active site.

To evaluate the catalytic potential of the newly introduced active site, we compared the number of esters being hydrolyzed with that of the wild-type protein ($EH1_A$), 144 naturally occurring ester hydrolases and two commercial preparations tested against the same set of 96 esters.[30]

As shown in Figure 6, we observed that $EH1_B$ had a restricted substrate spectrum (24 esters) compared to $EH1_A$ (72 esters), the most promiscuous ester hydrolase among those examined, but it was broader than that of the other hundred native ester hydrolases. The ester hydrolase with the new active site designed herein would fall into the category of ester hydrolases with moderate substrate promiscuity, thus suggesting the hydrolytic potential of the newly introduced active site in terms of the substrate spectrum.

We further evaluated whether $EH1_B$ showed different substrate preferences compared to $EH1_A$. We considered only those 24 substrates converted by both variants (Figure 2) and calculated the relative activity for each. As shown in Figure 2, distinct profiles were observed that highlight the difference in binding capacity for each of the catalytic environments. As an example, the new active site ($EH1_B$) showed a higher relative preference for the hydrolysis of methyl-, propyl-, and ethyl-hexanoate compared to the native site ($EH1_A$). The difference in substrate binding and preference between the new ($EH1_B$) and wild type ($EH1_A$) active sites was also noticeably when comparing their capacity to hydrolyze phenyl propionate and *p*NPP (Supporting Figure S9 and Table S3). Thus, in terms of k_{cat} $EH1_B$ preferred phenyl propionate over *p*NPP (5.7-fold), whereas $EH1_A$ slightly preferred *p*NPP (1.4-fold). A local refinement simulation with PELE revealed that the changes in preference for ethyl-hexanoate correlate with substrate placement. In particular, the difference in number of catalytic events in both sites, which define the poses where the substrate is readily placed for catalysis, is significantly reduced for ethyl-hexanoate, when compared to one of the substrates with

DOI: 10.1021/acs.biochem.8b00274
Biochemistry 2018, 57, 2245−2255

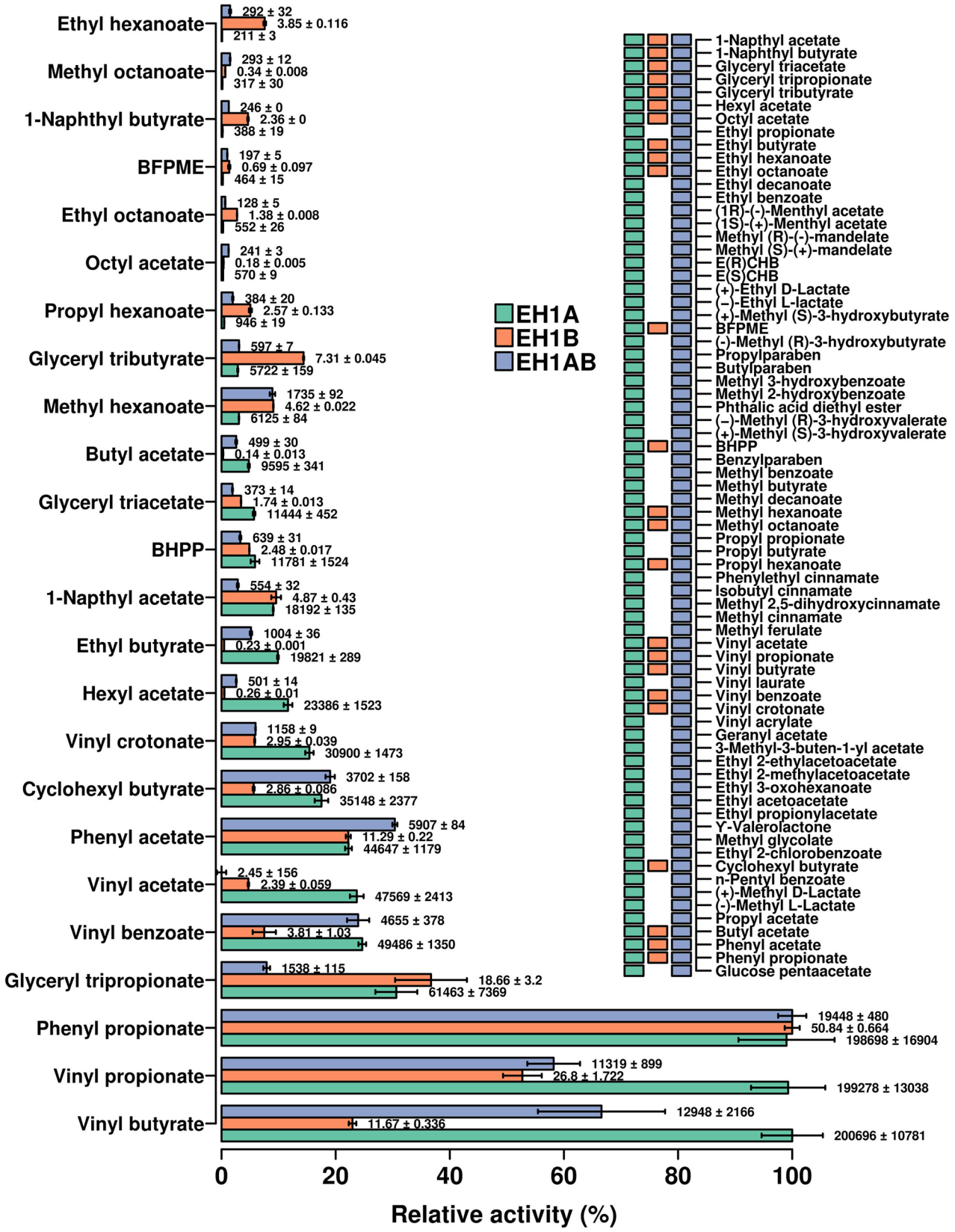


Figure 2. Substrate spectra of EH1 variants. The 72 out 96 esters for which activity was detected are listed on the right side panel, with indication of substrates being converted by each of the three variants. The ID code representing each EH variant is color coded. On the left main panel the relative activity (%) for each ester referred to the best substrate is indicated, with specific activity (U g^{-1}) and standard deviation (triplicates) indicated in the bars. Abbreviations as follows: E(R/S)CHB: ethyl (*R*/*S*)-(+)-4-chloro-3-hydroxybutyrate; BFPME: benzoic acid, 4-formyl-, phenylmethyl ester; BHPP: benzyl (*R*)-(+)-2-hydroxy-3-phenylpropionate. The activity protocol established and used to identify the esters hydrolyzed by each variant was based on a 550 nm follow-up pH indicator assay at pH 8.0 and 30 °C (see Experimental Section).

DOI: 10.1021/acs.biochem.8b00274
Biochemistry 2018, 57, 2245−2255

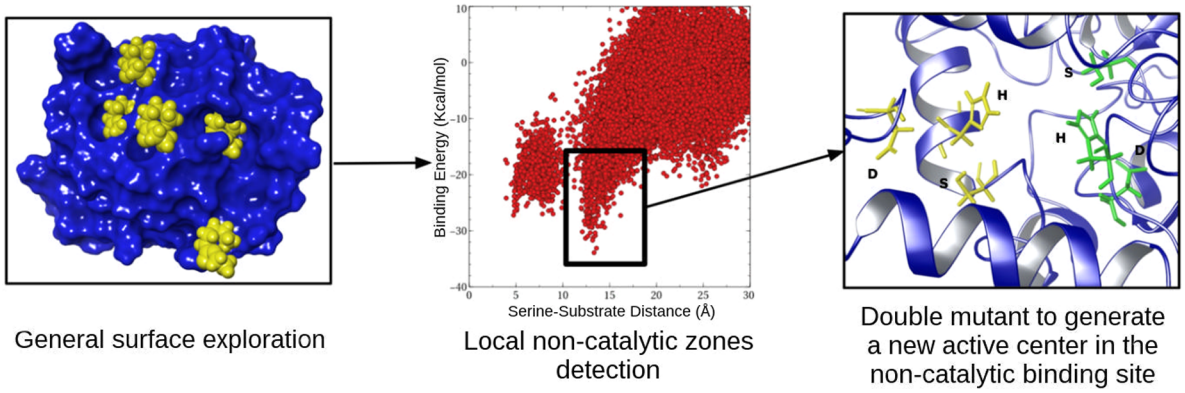


Figure 3. Molecular modeling protocol used for adding a second catalytic active site: (i) Complete surface exploration with the probe ligand (glyceryl tripropionate); (ii) alternative binding site location; (iii) substrate placement simulations (local refinement on the new site) with mutations.

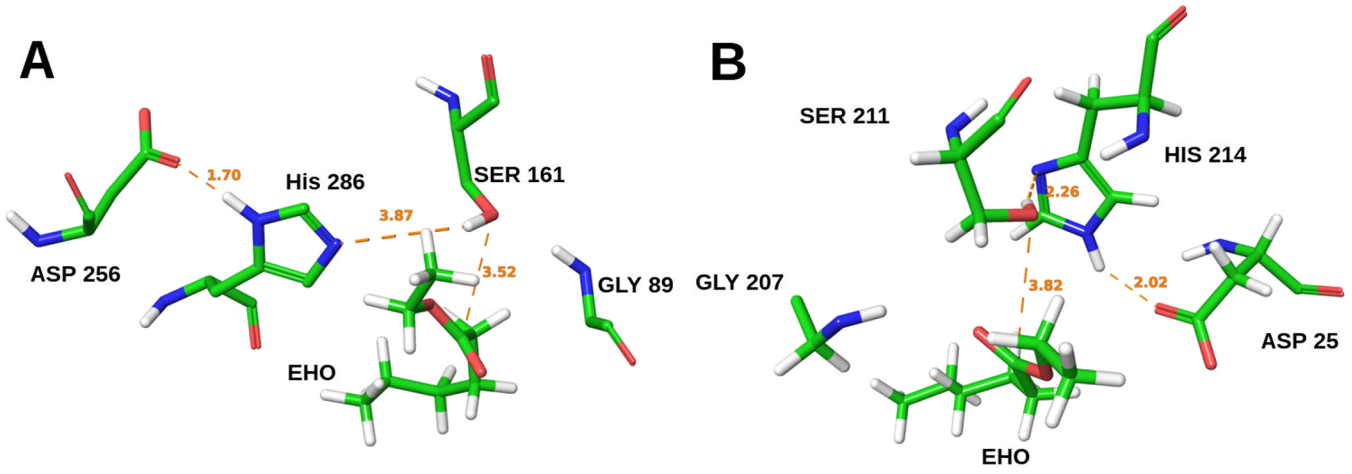


Figure 4. Catalytic position adopted by ethyl hexanoate (EHO) in the wild type active site (A) and in the Glu25Asp and Leu214His double mutant (B). Along with the substrate, the catalytic triad and the oxyanion hole residue are shown; main catalytic distances are underlined in angstroms.

maximum activity, phenyl propionate (Figure 7). Thus, differences in substrate profile may be due to distinct active site architectures (Figures 1 and 4). Additionally, there are different active site effective volumes (the active site cavity volume corrected by the relative solvent accessible surface area (SASA) of the catalytic triad) for the $EH1_B$ active site (82.2 Å^3) and the $EH1_A$ active site (167.7 Å^3). This volume has been recently correlated with changes in substrate promiscuity in ester hydrolases.[30]

Finally, we compared the hydrolytic rate of $EH1_B$ with that of $EH1_A$ and 225 previously reported ester hydrolases. We chose glyceryl tributyrate as a representative ester substrate because its specific activity values are commonly reported. As shown in Figure 2, the activity (7.3 ± 0.1 U g^{-1}) might seem moderate compared to that of $EH1_A$ (5722 ± 159 U g^{-1}). This was also proven by calculating the k_{cat}/K_M for phenyl propionate and *p*NPP, which was also found as a substrate (see Supporting Figure S9 and Table S3). On the basis of the modeled structure (Figure 4), the designed side chains were nearby but not quite in the typical orientation for the catalytic triad, which may explain why catalysis was less effective for $EH1_B$ compared to $EH1_A$. It is well-known that catalytic triads in serine ester hydrolases require a precise relative orientation of the Ser-Asp-His side chains.[11] Whatever the difference with $EH1_A$, we observed that the specific activity of $EH1_B$ is comparable to or higher than that of 93 previously reported native ester hydrolases that showed either low (≤6.8 U g^{-1}) or no (near zero) capacity to hydrolyze this substrate (about 70 esterases) (Figure 6 inset). The other reported 132 esterases were found far more active for the target substrate. Together, we can conclude that, at least for the hydrolysis of glyceryl tributyrate, the designed active site albeit being characterized by low catalytic activity, is as active as those of many natural hydrolases, but still not competitive with those of most esterases.

We found that single Ser211Ala, as well as Asp25Gln or His214Phe, substitutions in $EH1_B$ completely arrested the activity for all 96 chemically and structurally different esters examined in this study (Supporting Table S2), that included the 24 different esters hydrolyzed by $EH1_B$ (see Figure 2), and *p*NPP (see Supporting Results and Figures S9 and S10). This unambiguously confirms that they are the functional groups supporting catalysis in the new active site, which is constituted by a catalytic triad, and not a diad or a histidine. Altogether, these data demonstrated that it is plausible to use molecular modeling to find a potential extra binding pocket in a serine ester hydrolase that can be turned into a catalytic site. This site compares in terms of substrate spectra and catalytic activity with those from many other reported native ester hydrolases. It is difficult to make a statement whether the substrate spectrum of the newly introduced active site is comparable to, or is higher than that in the variants designed *de novo* by computational

DOI: 10.1021/acs.biochem.8b00274
Biochemistry 2018, 57, 2245−2255

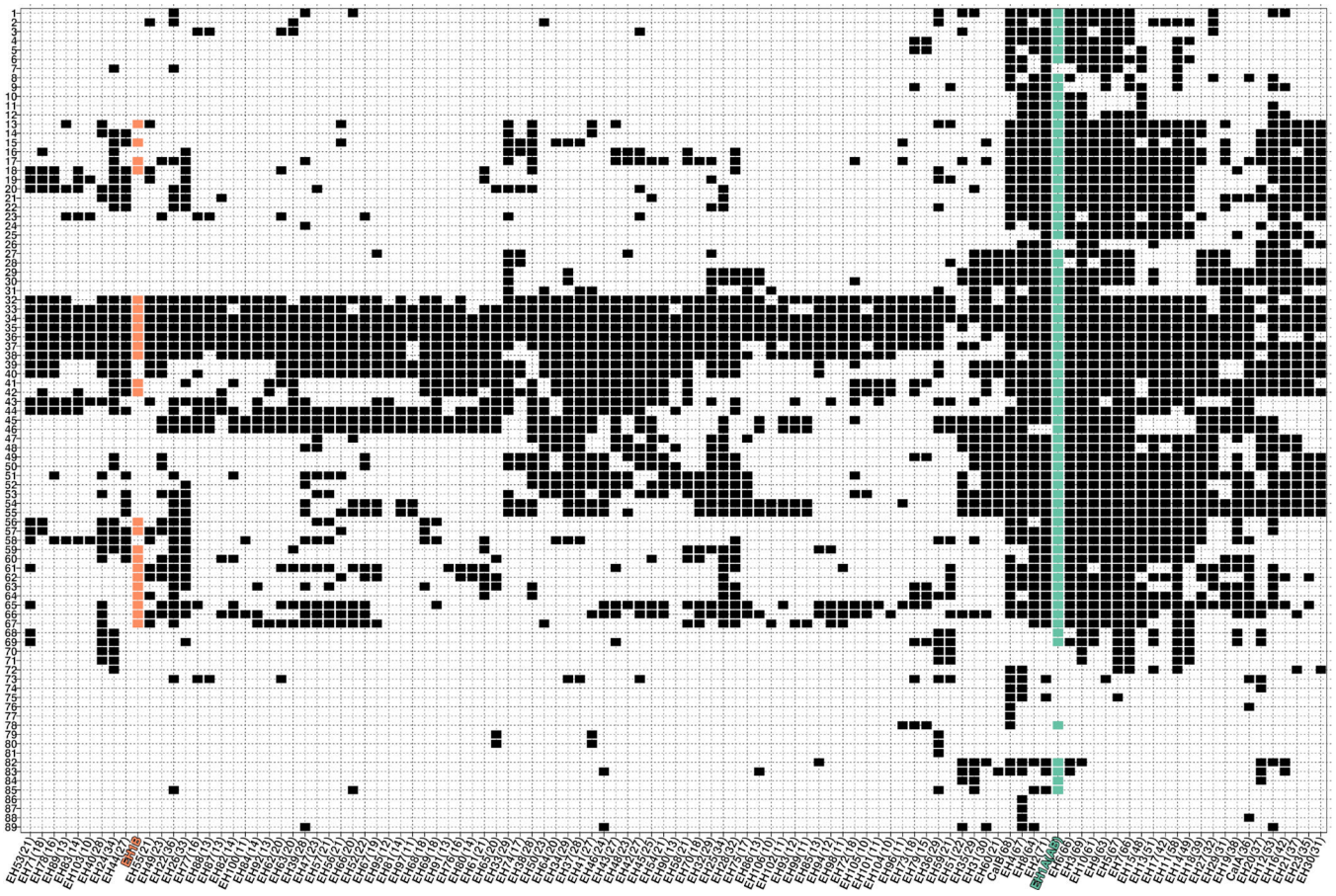


Figure 5. Clustering of the substrate range of EH1 variants within them and in comparison with those of other serine ester hydrolases. A total of 107 ester hydrolases, including the commercial preparations CalA and CalB from *Pseudozyma aphidis*, formerly *Candida antarctica*, are included. The data corresponding to each EH variant is color coded. This figure is created from data previously reported[30] and data herein obtained for EH1 variants (see Experimental Section). The list of the esters tested is shown on the left side (full name of the esters from no. 1 to 89 is given in Supporting Table S4). The ID code representing each ester-hydrolase is given at the bottom. Each hydrolase is named based on the code "EH", which means ester hydrolase, followed by an arbitrary number. The number in brackets indicates the number of esters hydrolyzed by each enzyme. The figure was created with the R language console, as described previously,[30] using a binomial table with information about the activity/inactivity (1/0) of the analyzed enzymes against the 96 substrates as a starting point.

tools previously described.[10−12,14,16] This is because these variants were only tested with a restricted set of *p*-nitrophenyl esters, substrates that are more easily hydrolyzed compared to the esters tested herein (Figure 2). It is clear though that, at least, $EH1_B$ is capable of *p*NPP hydrolysis with a k_{cat}/K_M of 16.4 M^{-1} s^{-1} (Supporting Figure S9 and Table S3), a value within the range or lower than that reported for other constructs using related substrates (k_{cat}/K_M from 18 to 309 M^{-1} s^{-1}).

Design and Characterization of a Serine Ester Hydrolase with Two Reactive Groups. An enzyme variant with the two active sites (the native and the newly identified) was created to prove that an enzyme with two hydrolytic active sites is functional. This mutant, named $EH1_{AB}$, was created by incorporating the Glu25Asp and Leu214His substitutions into the wild-type sequence. This two-active-site variant would likely employ Ser211 and Ser161 as nucleophiles (Figure 1). The corresponding gene was cloned, and the protein was expressed, purified (Supporting Figure S3 and Table S2), and characterized as described above.

The analysis of the variant $EH1_{AB}$ confirmed its ability to hydrolyze all 72 esters converted by $EH1_A$ (Figure 2). Thus, $EH1_{AB}$, as the native $EH1_A$, could be classified as an ester hydrolase with prominent substrate promiscuity (Figures 5 and 6). Because $EH1_A$, due to its promiscuous behavior, converted all esters hydrolyzed by $EH1_B$, we could not examine the effects of cooperativity for expanding the substrate spectra because of the presence of two different active sites in $EH1_{AB}$. We further observed that $EH1_{AB}$ does show overall a substrate preference similar to that of $EH1_A$ (Figure 2), although in some cases, the potential of the newly introduced active site to increasing the preference for a number of esters, such as methyl-, propyl-, and ethyl-hexanoate, was noticed. Also, $EH1_{AB}$ preferred phenyl propionate over *p*NPP (1.3-fold), whereas $EH1_A$ slightly preferred *p*NPP (1.4-fold) (Supporting Table S3). This may be a consequence of a higher preference of the second active site for phenyl propionate over *p*NPP.

The specific activity of $EH1_{AB}$ for all accepted esters was lower than that observed for $EH1_A$ but higher than that for $EH1_B$ (Figure 2). As example, using glyceryl tributyrate, its activity (597 ± 7 U g^{-1}) was 10-fold lower than that of $EH1_A$ and 81-fold higher than that of $EH1_B$. These differences were also observed when calculating the k_{cat}/K_M for phenyl propionate and *p*NPP (see Supporting Figure S9 and Table S3). The change in activity with respect $EH1_A$ could be explained by the decrease in substrate population for the main active site observed when introducing the mutations (Support-

DOI: 10.1021/acs.biochem.8b00274
Biochemistry 2018, 57, 2245−2255

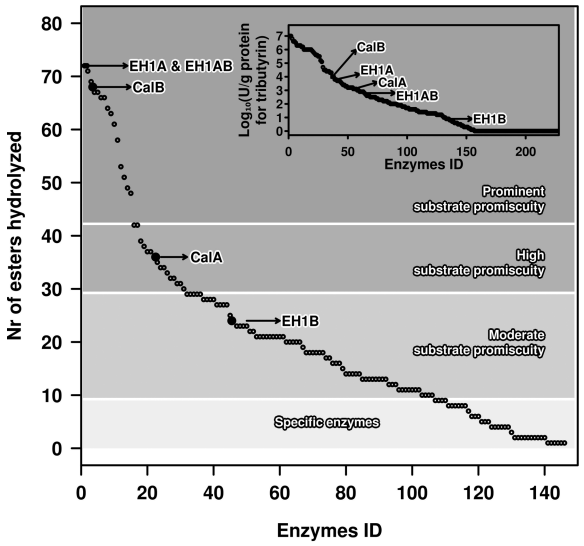


Figure 6. Comparison of the substrate range and catalytic performance of EH1 variants and other reported ester hydrolases. This figure is created from data previously reported[30] and data herein reported for EH1 variants (see Figures 2 and 5). The main figure represents the number of substrates hydrolyzed by each EH1 variant compared to that of other 146 serine ester hydrolases that included the commercial preparations CalA and CalB.[30] The inset represents the specific activity against glyceryl tributyrate of the EH1 variants and other 225 reported ester hydrolases (that included the 146 previously mentioned[30] and 79 reported elsewhere).

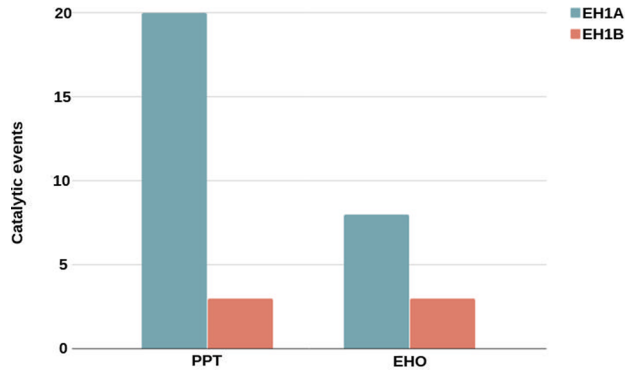


Figure 7. Number of catalytic events found in $EH1_A$ and $EH1_B$ catalytic sites along PELE simulations for ethyl hexanoate (EHO) and phenyl propionate (PPT). A catalytic event was defined when the main catalytic distances: ligand-Ser, Ser-His, and His-Asp, are lower than 3.5 Å. All simulations include the same sampling effort.

ing Figures S1a and b). Visual inspection of the models and energies (Supporting Figure S1b) shows a larger connection between the two sites, induced by changes in the hydrophobic packing (Leu24, Phe218, Leu24, and Trp217) after introducing His214. The α helix harboring the Leu214His mutation forms one face of the native active site, which slightly perturbs the main site. Whatever the catalytic activity level of the $EH1_{AB}$ variant compared to the native enzyme, its activity is significantly higher than that observed for 163 naturally occurring ester hydrolases (Figure 6, inset).

The consequences of the presence of two active sites were visible in K_M Michaelis–Menten (Supporting Figure S9 and Table S3) and pH optima (Supporting Figure S11) curves. Thus, the K_M progress curve for $EH1_{AB}$ displays a sigmoidal plot of the initial reaction rate versus substrate concentration (for both *p*NPP and phenyl propionate), rather than the hyperbolic plots observed for $EH1_A$ and $EH1_B$. This situation is typically found in allosteric enzymes, where the binding of substrates occurs in several active sites in the same enzyme molecule,[37] a situation that may most likely occurs in $EH1_{AB}$. We also observed that $EH1_{AB}$ displays an optimum pH profile distinct (narrower) to those of $EH1_A$ (which showed the broader range) and $EH1_B$ (see Supporting Results), which also differ between them (Supporting Figure S11). Subsequent investigation of the reasons that explain these differences will be needed.

We found that a combination of Ser211Ala/Ser161Ala, as well as Asp25Gln/Asp256Gln or His214Phe/His286Phe, substitutions in $EH1_{AB}$ completely arrested the activity of all 96 esters tested in this study, including the 72 esters converted by $EH1_{AB}$, as well as *p*NPP (see Supporting Results and Figures S9 and S10), unambiguously confirming their role in catalysis.

Altogether, these data unambiguously demonstrated that two different active sites can functionally and efficiently coexist in a single serine ester hydrolase. However, we did not identify a single specific advantage of this engineered ester hydrolase with two active sites, albeit we found that properties such as substrate binding and the optimum pH were altered. We hypothesized why the engineered enzyme had a lower performance. One of the mutations needed to introduce the new active site most likely slightly perturbs the native catalytic triad arrangement, which has negative consequences for the activity phenotype of the engineered mutant. Locating an extra binding pocket in positions not disturbing the original pocket and converting it into a catalytic site could solve this issue. In terms of substrate spectra, we should again highlight that $EH1_A$ already contains an active site capable of hydrolyzing a prominent number of esters, which is significantly higher than for the other 146 natural ester hydrolases, including the best commercial prototypes such as CalB (Figure 6).[30] Therefore, the probability of designing an active site capable of converting an additional set of esters compared to $EH1_A$ may be low. To unambiguously prove the advantage of introducing two active sites for expanding the substrate range, one should start with an ester hydrolase with a smaller substrate range and insert an extract active with different specificity or selectivity, which is actually being undertaken.

CONCLUSION

Is it possible to create enzymes with more than one active site? This idea, which has never been successfully attempted, was the starting point of the present study. Our work describes, for first time to the best of our knowledge, a neat proof-of-concept demonstration of the use of computational approaches to discover in an ester hydrolase already containing a native catalytic site, a novel binding site where a second catalytic site can be successfully introduced to generate a protein with two functional active sites for ester hydrolysis. Both active sites were structurally and catalytically different. We not only proved that introducing extra functional groups into a catalytic serine ester hydrolase is plausible, but also examined the catalytic potential of the resulting variant as compared to the target ester hydrolase and other reported native ester hydrolases. We confirmed the catalytic potential of the newly introduced active site as an individual catalytic entity or when combined with the native active site in the same polypeptide. We would like to highlight that a number of control experiments and analytics were performed to support our assertions and to ensure the

DOI: 10.1021/acs.biochem.8b00274
Biochemistry 2018, 57, 2245–2255

reported catalytic activity for the de novo designed active site was not due to a contaminant (for details see Supporting Results Section).

Can every ester hydrolase be evolved for multiple reactive groups? Possibly not, as introducing multiple catalytic environments would depend on the possibility of finding an extra binding site in an enzyme of a size large enough to bind into defined cavities. Also, if found, it may not be turned into a catalytic active site after introducing the appropriate mutations because catalysis requires amino acids at appropriate distances and angles.[11] Extending the analysis to other ester hydrolases and enzyme scaffolds will help to clarify whether the possibility to introduce extra catalytic sites is fortuitous or could be reproducible and to what extent can it happen.

The scientific and economic benefits of introducing multiple active sites into an enzyme are enormous and may revolutionize the fields of protein design and enzymology. Thus, converting binding pockets into extra catalytic active sites with different active site cavity volumes, accessible surface areas and distances, and angles between amino acids participating in the reactions compared to the native binding pockets will open novel catalytic opportunities for a given enzyme. These opportunities will come from the presence of different catalytic environments in the same polypeptide. Creating enzymes with multiple reactive groups may improve the substrate conversion range of enzymes otherwise containing only one active site, increase the competitiveness of enzymes over heterogeneous catalytic materials containing a high number of functional groups, and open application ranges with consequent reductions in production costs of other multiple enzymes. Moreover, adding new active sites to an enzyme whose original/native one cannot well be implemented for a particular application, but its overall structure has the appropriate technical requirements (scalability and robustness), may open new opportunities for a given enzyme. Such opportunities can be further explored through rational design and directed evolution focusing on residues around the new active site. Further studies in that direction would lay the foundation for streamlined and reproducible approaches for designing enzymes with multiple active sites. It would also promote new advances in a mechanistic understanding of enzymatic principles and the evolution of catalytic environments, particularly for revealing why certain areas in protein macromolecules are favored over others for incorporating active sites.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.8b00274.

Supporting Results Section, Supporting Figures S1−S11, and Supporting Tables S1−S4 (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*(V.G.) E-mail: victor.guallar@bsc.es.

*(M.F.) E-mail: mferrer@icp.csic.es.

ORCID

Gerard Santiago: 0000-0002-0506-3049

Peter N. Golyshin: 0000-0002-5433-0350

Víctor Guallar: 0000-0002-4580-1114

Manuel Ferrer: 0000-0003-4962-4714

Author Contributions

○G.S and M.M-M. contributed equally to this work.

Funding

This project received funding from the European Union's Horizon 2020 research and innovation program [Blue Growth: Unlocking the potential of Seas and Oceans] under Grant Agreement No. [634486] (project acronym INMARE). This research was also supported by the Grants PCIN-2014-107 (within ERA NET IB2 Grant No. ERA-IB-14-030 - MetaCat), PCIN-2017-078 (within the ERA-MarineBiotech grant Pro-Bone), BIO2014-54494-R, CTQ2016-79138-R, and BIO2017-85522-R from the Spanish Ministry of Economy, Industry and Competitiveness. P.N.G. gratefully acknowledges funding from the UK Biotechnology and Biological Sciences Research Council (Grant No. BB/M029085/1). R.B. and P.N.G. acknowledge the support of the Supercomputing Wales project, which is part-funded by the European Regional Development Fund (ERDF) via Welsh Government. P.N.G. acknowledges the support of the Centre of Environmental Biotechnology Project funded by the European Regional Development Fund (ERDF) through the Welsh Government. The authors gratefully acknowledge financial support provided by the European Regional Development Fund (ERDF). The MALDI-TOF/TOF analysis was performed in the proteomics facility of the Spanish National Center for Biotechnology (CNB-CSIC) that belongs to ProteoRed, PRB2-ISCIII, supported by Grant PT13/0001.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

C.C. thanks the Spanish Ministry of Economy, Industry and Competitiveness for a Ph.D. fellowship (Grant BES-2015-073829). The author would like to acknowledge Sergio Ciordia and Maria C. Mena for MALDI-TOF/TOF analysis.

■ REFERENCES

(1) Dufaud, V., and Davis, M. E. (2003) Design of heterogeneous catalysts via multiple active site positioning in organic-inorganic hybrid materials. *J. Am. Chem. Soc. 125*, 9403−9413.

(2) Jing, W., and DeAngelis, P. L. (2000) Dissection of the two transferase activities of the *Pasteurella multocida* hyaluronan synthase: two active sites exist in one polypeptide. *Glycobiology 10*, 883−889.

(3) Frickel, E. M., Jemth, P., Widersten, M., and Mannervik, B. (2001) Yeast glyoxalase I is a monomeric enzyme with two active sites. *J. Biol. Chem. 276*, 1845−1849.

(4) Deponte, M., Sturm, N., Mittler, S., Harner, M., Mack, H., and Becker, K. (2007) Allosteric coupling of two different functional active sites in monomeric *Plasmodium falciparum* glyoxalase I. *J. Biol. Chem. 282*, 28419−28430.

(5) Khare, S. D., Kipnis, Y., Greisen, P., Jr., Takeuchi, R., Ashani, Y., Goldsmith, M., Song, Y., Gallaher, J. L., Silman, I., Leader, H., Sussman, J. L., Stoddard, B. L., Tawfik, D. S., and Baker, D. (2012) Computational redesign of a mononuclear zinc metalloenzyme for organophosphate hydrolysis. *Nat. Chem. Biol. 8*, 294−300.

(6) Seelig, B., and Szostak, J. W. (2007) Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. *Nature 448*, 828−831.

(7) Toscano, M. D., Woycechowsky, K. J., and Hilvert, D. (2007) Minimalist active-site redesign: teaching old enzymes new tricks. *Angew. Chem., Int. Ed. 46*, 3212−3236.

(8) Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Rothlisberger, D., Zanghellini, A., Gallaher, J. L., Betker, J. L., Tanaka, F., Barbas, C. F., 3rd, Hilvert, D., Houk, K. N., Stoddard, B. L., and Baker, D. (2008)

DOI: 10.1021/acs.biochem.8b00274
Biochemistry 2018, 57, 2245−2255

De novo computational design of retro-aldol enzymes. *Science 319*, 1387−1391.
(9) Korendovych, I. V., Kulp, D. W., Wu, Y., Cheng, H., Roder, H., and DeGrado, W. F. (2011) Design of a switchable eliminase. *Proc. Natl. Acad. Sci. U. S. A. 108*, 6823−6827.
(10) Der, B. S., Edwards, D. R., and Kuhlman, B. (2012) Catalysis by a de novo zinc-mediated protein interface: implications for natural enzyme evolution and rational enzyme engineering. *Biochemistry 51*, 3933−3940.
(11) Richter, F., Blomberg, R., Khare, S. D., Kiss, G., Kuzin, A. P., Smith, A. J., Gallaher, J., Pianowski, Z., Helgeson, R. C., Grjasnow, A., Xiao, R., Seetharaman, J., Su, M., Vorobiev, S., Lew, S., Forouhar, F., Kornhaber, G. J., Hunt, J. F., Montelione, G. T., Tong, L., Houk, K. N., Hilvert, D., and Baker, D. (2012) Computational design of catalytic dyads and oxyanion holes for ester hydrolysis. *J. Am. Chem. Soc. 134*, 16197−16206.
(12) Zastrow, M. L., Peacock, A. F., Stuckey, J. A., and Pecoraro, V. L. (2012) Hydrolytic catalysis and structural stabilization in a designed metalloprotein. *Nat. Chem. 4*, 118−123.
(13) Korendovych, I. V., and DeGrado, W. F. (2014) Catalytic efficiency of designed catalytic proteins. *Curr. Opin. Struct. Biol. 27*, 113−121.
(14) Rufo, C. M., Moroz, Y. S., Moroz, O. V., Stohr, J., Smith, T. A., Hu, X., DeGrado, W. F., and Korendovych, I. V. (2014) Short peptides self-assemble to produce catalytic amyloids. *Nat. Chem. 6*, 303−309.
(15) Raymond, E. A., Mack, K. L., Yoon, J. H., Moroz, O. V., Moroz, Y. S., and Korendovych, I. V. (2015) Design of an allosterically regulated retroaldolase. *Protein Sci. 24*, 561−570.
(16) Moroz, Y. S., Dunston, T. T., Makhlynets, O. V., Moroz, O. V., Wu, Y., Yoon, J. H., Olsen, A. B., McLaughlin, J. M., Mack, K. L., Gosavi, P. M., van Nuland, N. A., and Korendovych, I. V. (2015) New tricks for old proteins: single mutations in a nonenzymatic protein give rise to various enzymatic activities. *J. Am. Chem. Soc. 137*, 14905−14911.
(17) Aranda, J., Cerqueira, N. M., Fernandes, P. A., Roca, M., Tunon, I., and Ramos, M. J. (2014) The catalytic mechanism of carboxylesterases: a computational study. *Biochemistry 53*, 5820−5829.
(18) Ferrer, M., Bargiela, R., Martínez-Martínez, M., Mir, J., Koch, R., Golyshina, O. V., and Golyshin, P. N. (2015) Biodiversity for biocatalysis: A review of the α/β-hydrolase fold superfamily of esterases-lipases discovered in metagenomes. *Biocatal. Biotransform. 33*, 235−249.
(19) Zastrow, M. L., and Pecoraro, V. L. (2013) Influence of active site location on catalytic activity in de novo-designed zinc metalloenzymes. *J. Am. Chem. Soc. 135*, 5895−5903.
(20) Ross, M. R., White, A. M., Yu, F., King, J. T., Pecoraro, V. L., and Kubarych, K. J. (2015) Histidine orientation modulates the structure and dynamics of a de novo metalloenzyme active site. *J. Am. Chem. Soc. 137*, 10164−10176.
(21) Borrelli, K. W., Cossins, B., and Guallar, V. (2010) Exploring hierarchical refinement techniques for induced fit docking with protein and ligand flexibility. *J. Comput. Chem. 31*, 1224−1235.
(22) Hernandez-Ortega, A., Borrelli, K., Ferreira, P., Medina, M., Martínez, A. T., and Guallar, V. (2011) Substrate diffusion and oxidation in GMC oxidoreductases: an experimental and computational study on fungal aryl-alcohol oxidase. *Biochem. J. 436*, 341−350.
(23) Carlson, H. A., Smith, R. D., Damm-Ganamet, K. L., Stuckey, J. A., Ahmed, A., Convery, M. A., Somers, D. O., Kranz, M., Elkins, P. A., Cui, G., Peishoff, C. E., Lambert, M. H., and Dunbar, J. B., Jr. (2016) CSAR 2014: a benchmark exercise using unpublished data from pharma. *J. Chem. Inf. Model. 56*, 1063−1077.
(24) Santiago, G., de Salas, F., Lucas, M. F., Monza, E., Acebes, S., Martínez, A. T., Camarero, S., and Guallar, V. (2016) Computer-aided laccase engineering: toward biological oxidation of arylamines. *ACS Catal. 6*, 5415−5423.
(25) Lecina, D., Gilabert, J. F., and Guallar, V. (2017) Adaptive simulations, towards interactive protein-ligand modeling. *Sci. Rep. 7*, 8466.
(26) Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., and Jorgensen, W. L. (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B 105*, 6474−6487.
(27) Bochevarov, A. D., Harder, E., Hughes, T. F., Greenwood, J. R., Braden, D. A., Philipp, D. M., Rinaldo, D., Halls, M. D., Zhang, J., and Friesner, R. A. (2013) Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *Int. J. Quantum Chem. 113*, 2110−2142.
(28) Shan, Y., Kim, E. T., Eastwood, M. P., Dror, R. O., Seeliger, M. A., and Shaw, D. E. (2011) How does a drug molecule find its target binding site? *J. Am. Chem. Soc. 133*, 9181−9183.
(29) Martínez-Martínez, M., Alcaide, M., Tchigvintsev, A., Reva, O., Polaina, J., Bargiela, R., Guazzaroni, M. E., Chicote, A., Canet, A., Valero, F., Rico Eguizabal, E., Guerrero, M. C., Yakunin, A. F., and Ferrer, M. (2013) Biochemical diversity of carboxyl esterases and lipases from Lake Arreo (Spain): a metagenomic approach. *Appl. Environ. Microbiol. 79*, 3553−3562.
(30) Martínez-Martínez, M., Coscolín, C., Santiago, G., Chow, J., Stogios, P., Bargiela, R., Gertler, C., Navarro-Fernandez, J., Bollinger, A., Thies, S., Mendez-García, C., Popovic, A., Brown, G., Chernikova, T. N., García-Moyano, A., Bjerga, G. E. K., Perez-García, P., Hai, T., Del Pozo, M. V., Stokke, R., Steen, I. H., Cui, H., Xu, X., Nocek, B., Alcaide, M., Distaso, M., Mesa, V., Pelaez, A. I., Sanchez, J., Buchholz, P. C. F., Pleiss, J., Fernandez-Guerra, A. F., Glockner, F. O., Golyshina, O. V., Yakimov, M. M., Savchenko, A., Jaeger, K.-E., Yakunin, A. F., Streit, W. R., Golyshin, P. N., Guallar, V., Ferrer, M., and The INMARE Consortium (2018) Determinants and prediction of esterase substrate promiscuity patterns. *ACS Chem. Biol. 13*, 225−234.
(31) Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature 227*, 680−685.
(32) Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem. 72*, 248−254.
(33) Janes, L. E., Lowendahl, C., and Kazlauskas, R. J. (1998) Quantitative screening of hydrolase libraries using pH indicators: Identifying active and enantioselective hydrolases. *Chem. - Eur. J. 4*, 2324−2331.
(34) Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013) Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput.-Aided Mol. Des. 27*, 221−234.
(35) Shivakumar, D., Williams, J., Wu, Y., Damm, W., Shelley, J., and Sherman, W. J. (2010) Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J. Chem. Theory Comput. 6*, 1509−1519.
(36) Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem. 68*, 850−858.
(37) Berg, J. M., Tymoczko, J. L., and Stryer, L. (2002) *Biochemistry* 5th ed., W. H. Freeman, New York; Section 8.4. The Michaelis-Menten model accounts for the kinetic properties of many enzymes. Available from https://www.ncbi.nlm.nih.gov/books/NBK22430/.

DOI: 10.1021/acs.biochem.8b00274
Biochemistry 2018, 57, 2245−2255

This Thesis is aware that the enzymes that are commercially available are limited and not tailored to convert industrially relevant molecules under commonly used conditions in industrial processes. Exploring naturally occurring enzymes that have evolved to convert molecules of relevance is thus of high interest. We anticipated that, in order to identify and improve the value of new enzymes for industry, one needs to comprehensively compare their properties to find biocatalytic patterns that allow predicting those in databases. For that, it is important to cover an extensive enzyme diversity, so that this feature to be investigated and characterized is not dominated by a particular type of protein or highly similar clusters but consists of diverse non-redundant sequences assigned to multiple folds and sub-families, which are distantly related to known homologs in many cases, and originated from as much geographically distinct sites and microbial taxa as possible. This Thesis is also aware that in order to identify enzymes with properties of promiscuity, specificity, selectivity, and stability superior to those of the best commercial prototypes, one needs to extend the analysis to a broad range of conditions and substrates, including those of industrial importance. Also, applying different immobilization and engineering methods to the newly discovered enzymes is an effective strategy to help improving further their catalytic performance and stability or to give rise to new natural or non-natural reactions or specificities. By doing so, one can find and design enzymatic preparations attractive for industry in the end.

Following on from this, the industrial growing requirements in sustainable chemistry drive the search of new biocatalysts capable of performing a large variety of conversions. For this reason, in this Thesis, we first focused on enzymatic substrate promiscuity, which characterizes the range of substrates that react with a biocatalyst (**Figure 15**).

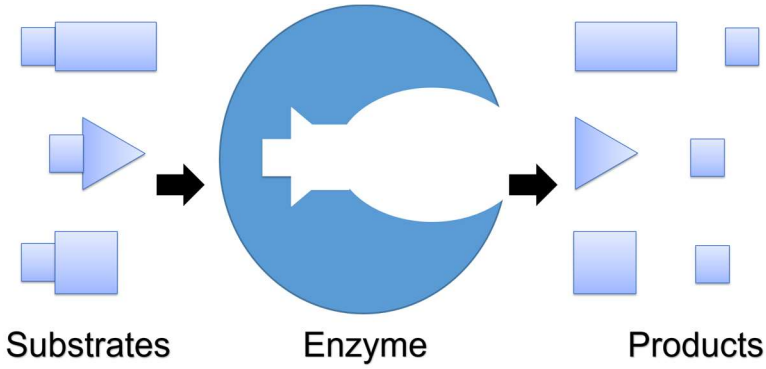


Figure 15. Schematic representation of the capacity of enzymes to convert and produce different chemicals while catalyzing the same reaction.

Deciphering a substrate promiscuous enzyme, however, requires a huge amount of experimental tests and it is really difficult to obtain in an accurate way. For this reason, we have contributed to creating one of the largest enzyme collections worldwide in a single laboratory and tested it with a broad range of substrates. By using a screening tree concept (**Figure 16**), the enzymes are pre-selected on the basis of their activity with few commonly used or industrially not-relevant substrates (i.e., 1-naphthyl acetate and glyceryl tributyrate in case of ester-hydrolases, and benzaldehyde and 2-(4-nitrophenyl)ethan-1-amine and o-xylylendiamine in case of transaminases); in some cases, new and sensitive methods have been adapted (Reyes-Duarte et al., 2018; Rodrigues, Sanches, & de Carvalho, 2019). Once the activity is confirmed, each enzyme is subjected to an extensive substrate fingerprint to evaluate the presence of promiscuous phenotypes. In this Thesis, we tested a broad range of substrates, which accounted at least 132 chemically and structurally diverse compounds (at least 96 used to screen ester-hydrolases and at least 36 in case of transaminases).

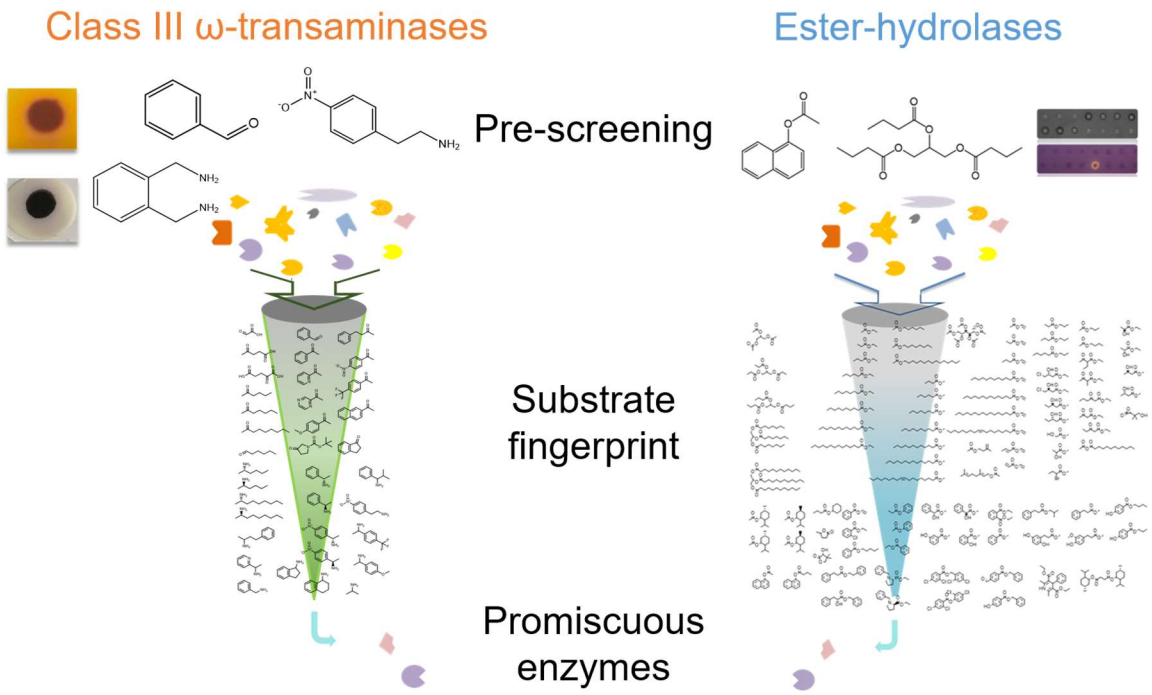


Figure 16. Scheme of the screening tree concept for searching substrate promiscuous enzymes. A high number of enzyme candidates is reduced to a small group of promiscuous ones through a pre-screening with common substrates and a further extensive screening with a diverse substrate library. For a detailed view of the chemical structure of substrates see **Figure 18** and **Figure 19**.

Based on the pipeline summarized in **Figure 16**, this Thesis has accumulated a number of quantifiable outcomes, that are detailed below. First, the creation of one of the largest collections of enzymes available to date in a single laboratory, most covering activities with primary use in bio-catalysis and industrial processes, namely, ester-hydrolases and class III ω-TAs. From this collection, extensive bio-catalytic, structural and computational information has been generated on the basis of which we have found a number of enzyme targets capable of manufacturing bio-based commodities to an extent higher than commercial products.

The number of identified genes encoding enzymes of interest, ester-hydrolases and transaminases, at the end of the Thesis accounts 155, from which 145 were ester-hydrolases and 10 class III ω-TAs, selected for comparative purposes and with validated activities. A sub-set of 20 derived from genome mining search (*Pseudomonas oleovorans, Staphylococcus epidermis, Geobacillus sp., Alcanivorax borkumensis, Oleiphilus messiensis*), and 135 from metagenome mining search from microbial communities of at least 28 environmental different sites; 110 of the latter were identified by naïve screenings in clone libraries and 25 after *in silico* mining from sequences derived from metagenomes. A schematic representation of the number of sequences found by each screening method is shown in Figure 17.

The 155-Enzyme Collection consists on highly diverse non-redundant sequences assigned to multiple folds and sub-families (at least 8 in case of the ester-hydrolases, and 1 in case of the transaminases), which are from 29.1 to 99.9% identical to known homologs in databases. The average pairwise amino acid sequence identity among all sequences of each of the two classes investigated range from 0.2 to 99.7%, indicating the high diversity covered. The 155-Enzyme Collection was also shown to be highly diverse and unique from a taxonomic point of view as it was originated from at least 50 known bacterial and archaeal genera. **Figure 17** illustrates the sites from which enzymes investigated in this Thesis were recovered.

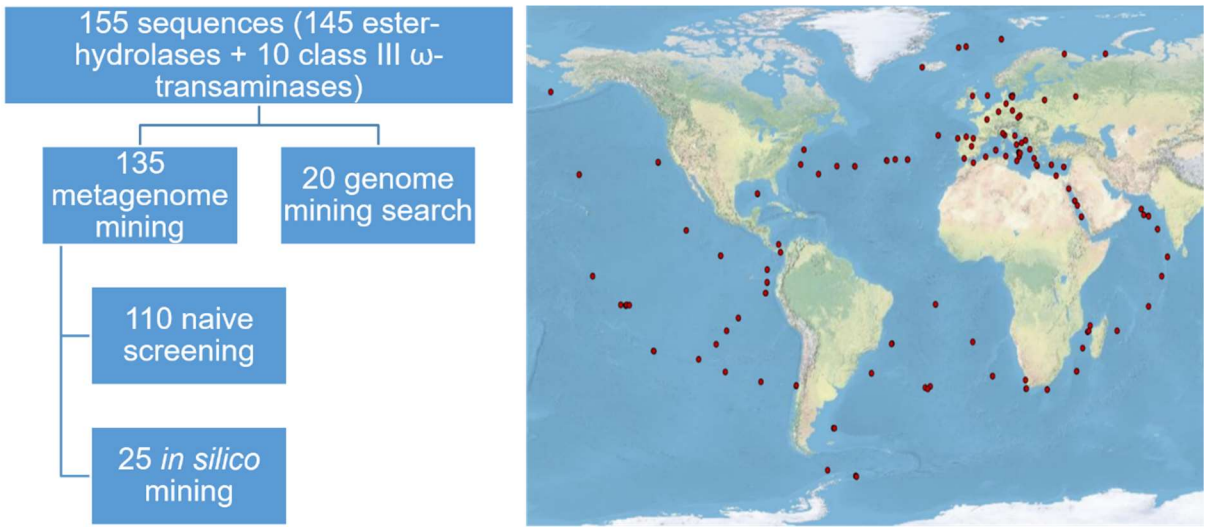


Figure 17. Map illustrating sites from where Thesis´s enzymes have been recovered with a classification graph of the number of sequences found by each screening method and source. The 28 different environmental sites from which the 155 enzymes were characterized are shown together with the 242 TARA Oceans samples from 5 oceanic regions from which presumptive promiscuous ester-hydrolases were selected by *in silico* search as described in Supplementary Information of Chapter 1.

At least 11 different types of vectors (pET-46 Ek/LIC, p15Tv-L, pBXCH, pBXNH3, pBXNHS, pCDFDuet, pCR-XL-TOPO, pET21a, pET22b, pVLT31 and pRhokHi-2) and 5 different host that included different strains of the bacteria *Escherichia coli* (strains BL22, BL21(DE3), MC1061, Top10 and DH5α) were used through this Thesis to clone, express and produce all 155 enzymes.

All 155 genes encoding enzymes of interest have been successfully cloned and expressed in soluble and functional form. Small scale fermentation tests were undertaken to find optimal conditions for expression. Also, all enzymes were purified by 6xHis affinity tag at >98% purity and extensively characterized.

Enzymes have been purified mostly by immobilized metal affinity chromatography and characterized by multiple methods that include colorimetric methods, chromatographic methods, mass spectrometry and crystallographic methods. A broad range of at least 132 chemically and structurally diverse compounds (including some of industrial interest), and operational conditions that included pH from 4.0 to 10.0, temperatures from 4 to 70˚C and the presence of at least 10 different solvents of industrial relevance (from 0 to 90% v/v) have been considered in the tests. **Figure 18** and **Figure 19** exemplified the chemical structure of substrates used for substrate profiling of ester-hydrolases and transaminases.

A set of 25 enzymes has been selected for potential application purposes based on their activity and stability compared to commercial products.

The selected enzymes include 10 amine transaminases of the class III ω-TAs capable of accepting best large and aromatic ketones, aldehydes and amines (see Chapter 3), and 15 ester-hydrolases (see Chapter 1) with a substrate range above those of commercial products (i.e. CalA and CalB).

Two crystal structures of ester-hydrolases have been solved and homology models were created from at least 104 enzymes (94 ester-hydrolases and 10 transaminases).

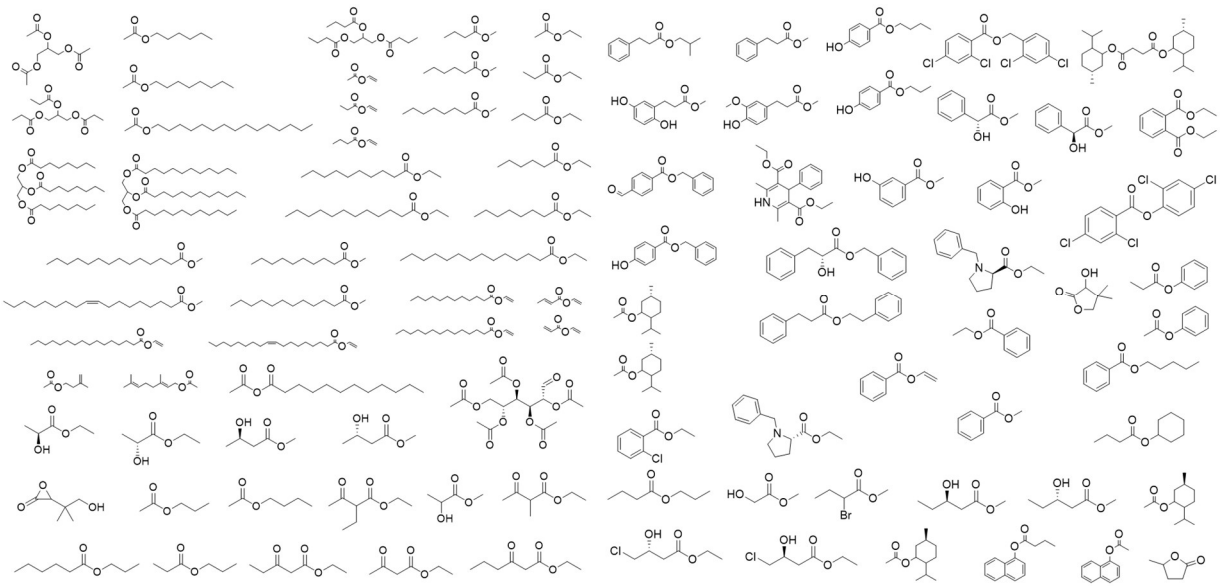


Figure 18. Chemical structure of representative esters used for profiling the substrate specificity of ester-hydrolases. For extensive description see Chapter 1.

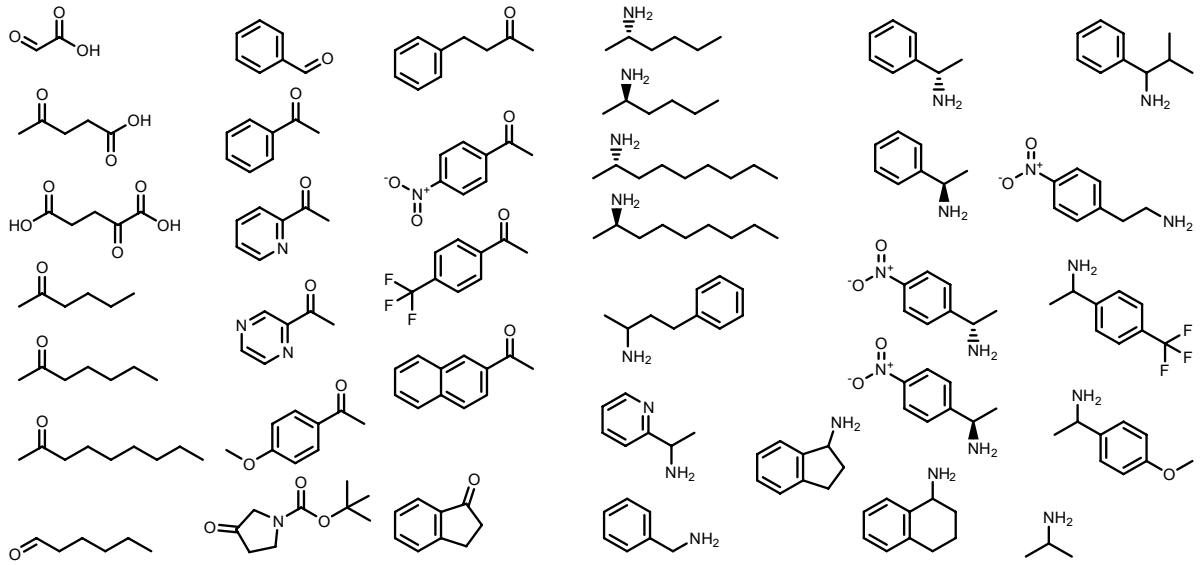


Figure 19. Chemical structure of representative ketones, aldehydes and amines used for profiling the substrate specificity of class III ω-TAs. For extensive description see Chapter 3.

The structures and models have been obtained with a double aim. First, to help understanding enzyme properties of promiscuity and selectivity. Second, to help applying computational and rational engineering tools to generate mutants with improved properties.

A set of 4 different immobilized preparations, from the top-selected ester-hydrolase based on its broad substrate specificity, was prepared (using at least 3 different techniques) and evaluated for small-scale biocatalysis tests, including 6 pairs of chiral esters.

In detail, target enzymes have been immobilized in three different supports, including inside the pores of amino-functionalized ordered mesoporous material with two different pore sizes, and on the surface of two types of magnetic micro-particles with different chemistry of linking. The activity tests revealed that

it is possible to create from a single enzyme different stereo-specific biocatalysts for a variety of chiral compounds, depending on the immobilization technique applied and the flexibility constraints associated with those.

Finally, at least 4 engineered mutants of an ester-hydrolase have been designed, that represent the first examples of a PluriZyme, an enzyme with two active sites.

In detail, we have shown that it is possible to add a second hydrolytic active site to an ester-hydrolase and that through this it is possible to intensify the bio-catalytic properties of the original enzyme. In the particular case of the ester-hydrolase used as a model, we found it could potentially increase the substrate range because of the synergetic effect of both the native and artificial sites, but also potentially introduce stereo-specificity (%e.e. >99.9%).

All the above information provided quantifiable outcomes of this Thesis. **Figure 20** summarizes the work pipeline used to generate such outcomes, including the enzymes, structures, 3D models, mutants and products.

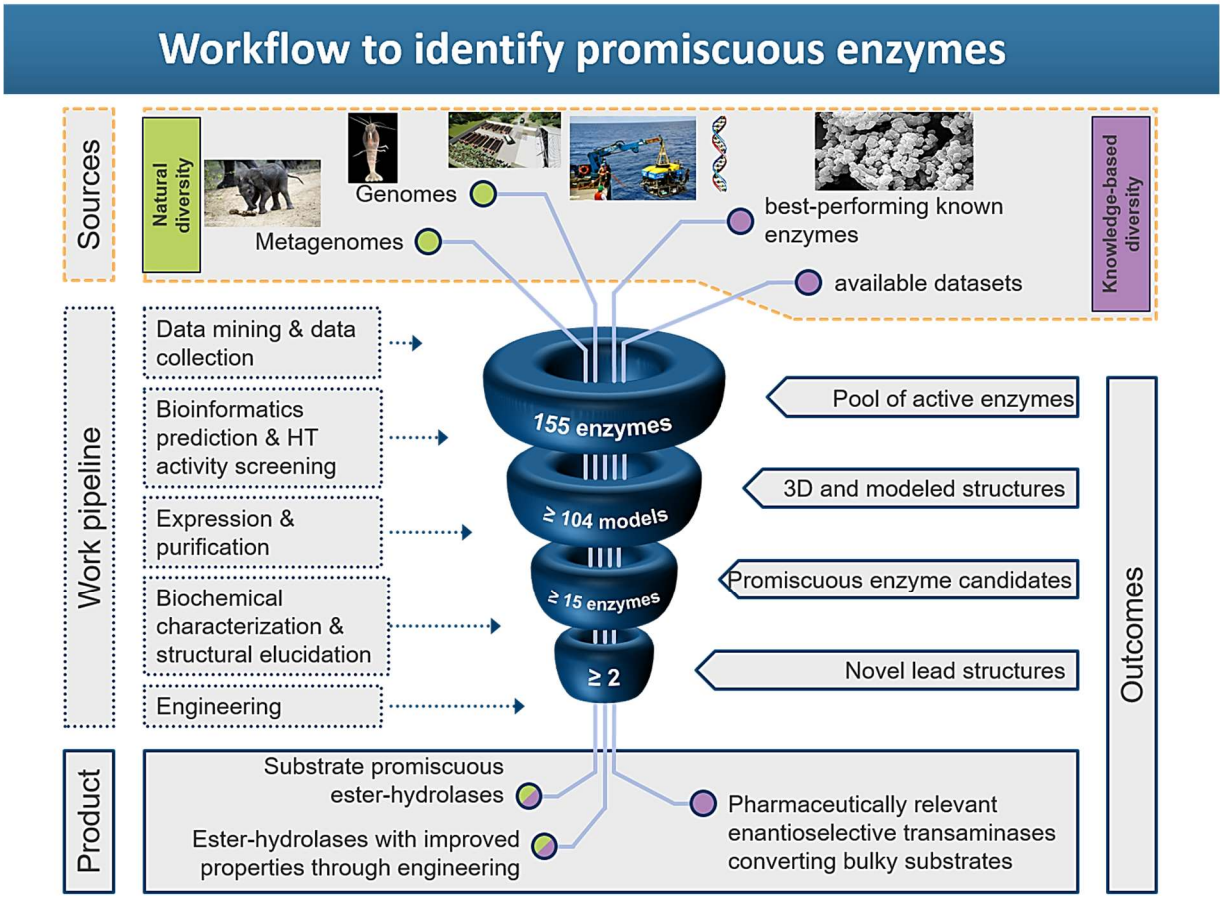


Figure 20. Methodological workflow applied in this Thesis and the quantifiable outcomes obtained within the pipeline.

Aside from the above quantifiable outcomes, a number of major outcomes obtained after integration of generated multi-data are worth mentioning, which are detailed and discussed below.

First, and most important, concerning the understanding of mechanisms underlying promiscuity and selectivity, we have found those mechanisms for two important classes of enzymes in industry: ester-hydrolases and class III ω-TAs.

More in detail, we found for the first time a structural parameter that helps classifying promiscuity level of ester-hydrolases from sequence data at 94% accuracy. This parameter, the active site effective volume, exemplifies the topology of the catalytic environment by measuring the active site cavity volume corrected by the relative solvent accessible surface area (SASA) of the catalytic triad; sequences encoding serine ester-hydrolases with active site effective volume (cavity volume/SASA) above a threshold show greater substrate spectra. Thus, based on the determination and utilization of this predictive marker, easily calculable from sequence data, anyone can fast-track the identification of industrially versatile/promiscuous ester-hydrolases without the need of tedious cloning and expression; just the sequence information is enough (**Figure 21**).

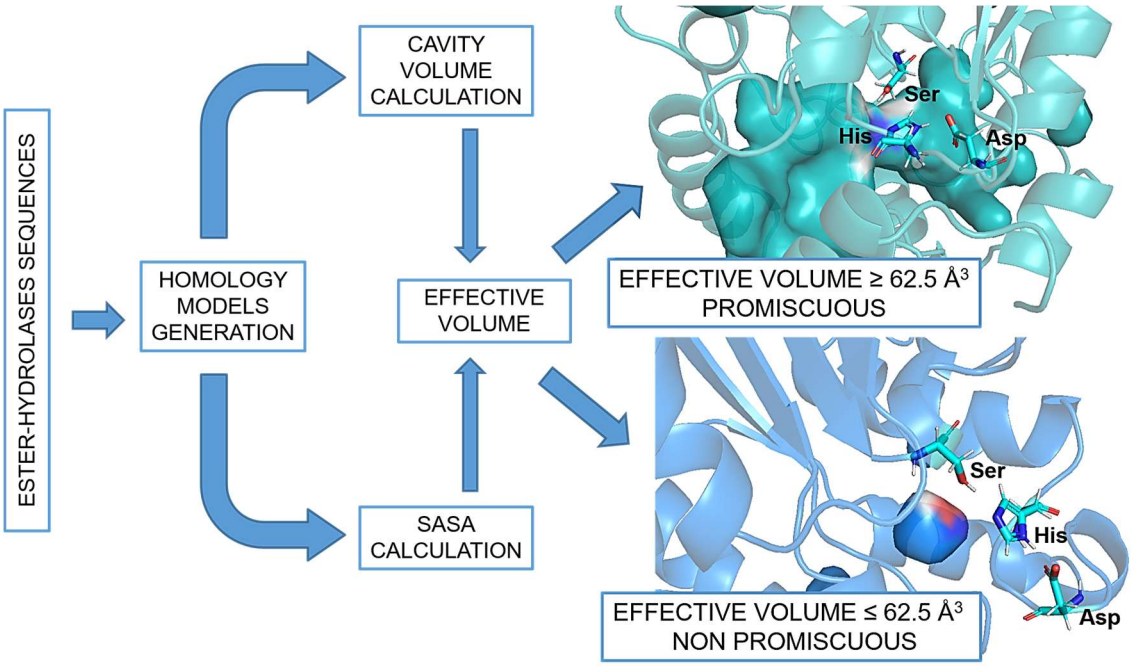


Figure 21. Pipeline to select sequences encoding ester-hydrolases with high cavity volume (shown as surface) and low SASA or exposure of the catalytic triad (shown as sticks) through databases. The ratio among those parameters determines, with high probability, the level of substrate promiscuity.

Whether the predictive marker for "substrate promiscuity" in ester-hydrolases can or cannnot be extrapolated to other enzyme families is yet to be established, but the demonstration in Chapter 1 that it could be extended to other enzyme classes, such as phosphatases of the haloacid dehalogenase (HAD) superfamily of hydrolases, suggests it is a possibility, at least, in hydrolases. In the case of class III ω-TAs, we found it was not the case. This was shown by characterizing a number of the aforesaid enzymes and by examining the molecular basis of substrate promiscuity. We found that the effective volume was not associated with the possibility of class III ω-TAs to convert more or less substrates, including bulky ones. Rather, we deciphered for the first time the reasons why the class III ω-TAs are capable of converting bulky substrates as well as (R)-amines, a desired feature from an industrial point of view. In detail, we demonstrated that the preference for bulky ketones was associated with the presence of a hairpin region proximal to the conserved flipping arginine (Arg) and residues close to it. The outward orientation of this Arg additionally favored the conversion of (R)-amines. Thus, just by checking a sequence presumptively encoding a class III ω-TAs for the presence of a hairpin region proximal to the conserved Arg, and by checking a homology model for the orientation of this conserved residue, anyone can speed up the identification of industrially promiscuous class III ω-TAs capable of converting bulky ketones, aldehydes and (R)-amines, without the need of time-consuming cloning and expression; just the sequence or model information is enough. Together, by following the pipeline summarized in **Figure 22**, one can suggest from

sequence information the possibility that it encodes a class III ω-TAs being capable of converting or not bulky substrates and (R)-amines, an uncommon feature among this class of enzymes.

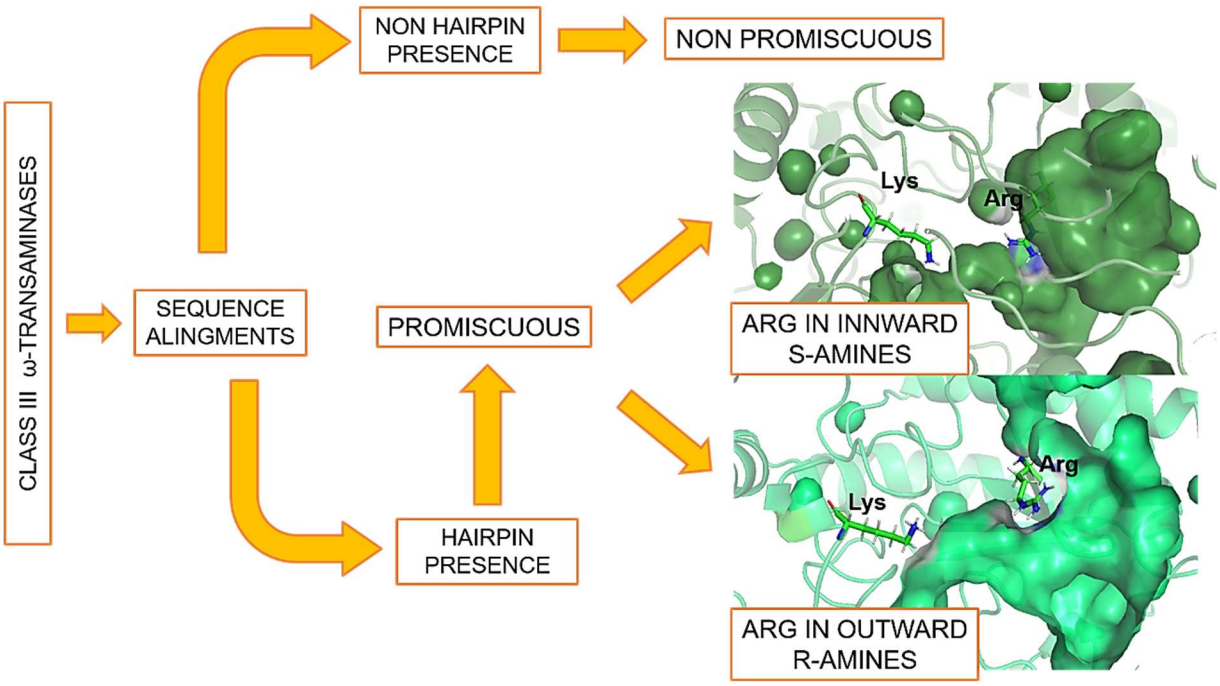


Figure 22. Pipeline to select in databases sequences encoding class III ω-TAs with high probability to convert bulky substrates and (R)-amines.

Together, the integration of bio-catalytic data, in combination with bioinformatic and computational tools, may help understanding by just using sequence or homology model information whether an ester-hydrolase does have a broad substrate specificity or not and whether a class III ω-TAs could convert bulky substrates and chiral amines. Having said that, while in the case of transaminases we could identify the molecular basis of substrate promiscuity, capacity to convert bulky substrates and enantio-selectivity, this was not the case of the ester-hydrolases, for which only a marker of promiscuity level could be unambiguously obtained. No marker of enantio-selectivity could be identified, and thus this property may vary from case to case.

The results presented demonstrate that the level of substrate promiscuity can be inferred from sequence and/or structure information. A step further will be to specifically determine which amino acids play a role in tuning the specificity during evolution. This will help us understand the loss or acquisition of capabilities to convert a specific set of substrates based on the loss or acquisition of specific amino acids; their identification may, in turn, also help to predict substrate specificities. In this direction, in the frame of a short stay at the laboratory of professor Olivier Lichtarge in the Baylor College of Medicine, we have explored the possibility of predicting the substrate specificity of ester-hydrolases and the amino acids implicated on it through the application of a software program called Evolutionary Trace (http://lichtargelab.org/software/ETserver). This software can score the amino acids of a protein attending at its conservation through an alignment of homolog sequences. Starting from this, we selected, among all sequences encoding ester-hydrolases, those belonging to family IV from Arpigny & Jaeger (1999). The reason was that we found, as described in Chapter 1, that this family contains the ester-hydrolases with broader substrate spectra. We performed an Evolutionary Trace. Finally, we merged the score ranking of amino acids with the substrate profile, grouping the substrates by chemical nature. We found a region from the cavity entrance of most Family IV members that, according to the trace, has some residues that are most likely implicated in the recognition of certain types of chemicals. A scheme of the

region position at the cavity entrance and the residues selected is shown in **Figure 23.** Currently, this region is being explored more in detail in order to clarify its implication in the evolution of substrate specificity within Family IV ester-hydrolases. Having said that, we should stress that investigating substrate specificity and reaction rates is no easy task, as we know that multiple local and global factors affect the capacity of an enzyme to convert a molecule. Also, it has been demonstrated that the conversion rate for a substrate does not only depend on its positioning in the active site but also on the reorganization of the electric fields (Welborn, Ruiz Pestana, & Head-Gordon, 2018). Thus, multiple factors need to be considered in the future.

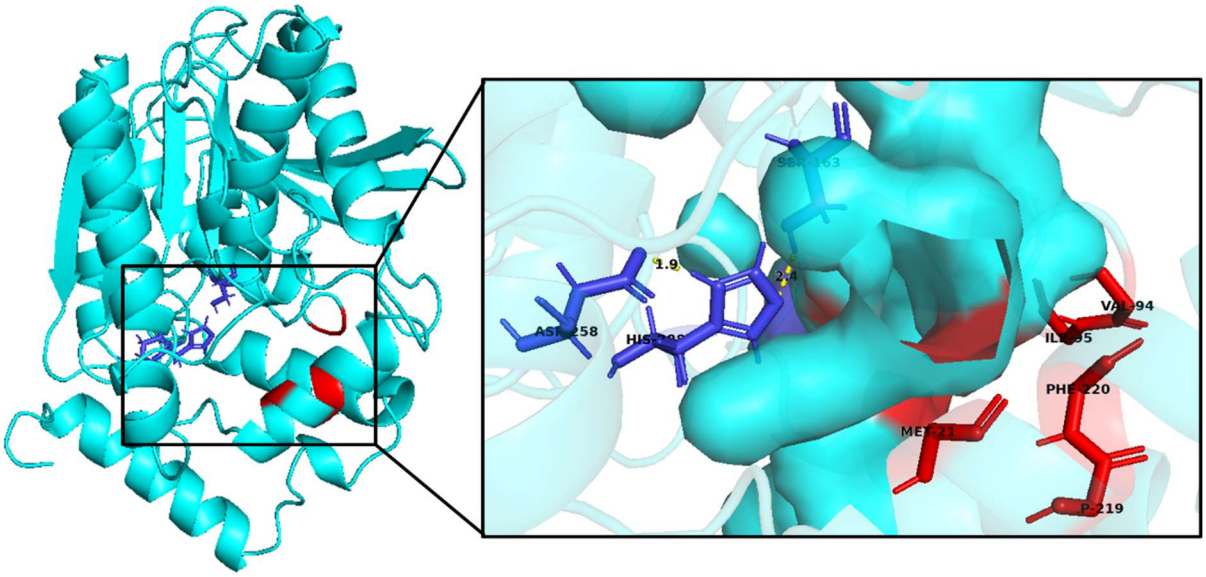


Figure 23. Scheme of EH1 serine ester-hydrolase (PDB 5JD4) with catalytic triad shown as dark blue sticks and putative substrate specificity modulator region shown in red in touch with catalytic cavity surface.

Second, this Thesis generates information suggesting that particular groups of microbes and environments may be potentially enriched in promiscuous ester-hydrolases and transaminases.

The possibility to associate each of the 155 enzymes recovered and investigated to different taxonomic groups helped us to evaluate whether or not certain microbes do contain enzymes more promiscuous than others. The data presented in Chapters 1 and 3 suggest that this is plausible. Indeed, the results presented in this Thesis suggest that microbes associated to *Archaeglobus, Methanomicrobiales, Staphylococcus, Geobacillus, Bacillus, Hloplasma, Symbiobacterium, Ruminoccaceae, Acidimicrobium, Ferrimicrobium, Adicitrix, Thermus, Serratia, Halioglobus, Idiomarina, Alteromonas, Alcanivorax, Oleiphilus, Marinobacter, Pseudomonas, Azotobacter, Cycloclasticus, Thiotrix, Variovorax, Cupriavidus, Ralstonia, Burkholderia, Immunidisolibacter, Nitrobacter, Hyphomonas, Rhizobium, Mesorhizobium, Rhodobacterae, Sphingomonas, Aciphilum, Azospirilum, Amphritea and Acidihalobacter* are containers of promiscuous enzymes to an extent higher than others. This was shown by finding that enzymes associated with these taxonomic groups have a broader substrate specificity than enzymes associated with other microbes.

Third, this Thesis was able to provide an estimation of the percentage of promiscuous enzymes in genomes and metagenomes and its relation to stereo-selectivity/specificity.

In particular, the presented data revealed that the probability of finding a "substrate promiscuous" ester-hydrolase (able to convert more than 50% of all tested chemicals, 96 in cases of ester-hydrolases) in meta-genomes is about 13%. That is, 13 out of 100 ester-hydrolases in genomes and metagenomes may show

a high-to-prominent promiscuity level. We also found: 1) the probability of finding ester-hydrolases with broad active site environment to accommodate very large aromatic substrates is relatively high (up to 48%); 2) no correlation exists between "substrate promiscuity level" and the "activity level", so that enzymes with a large active site capable of accepting a large number of esters may not be necessarily the most active ones; and 3) although 23% of all ester-hydrolases do show prominent enantio-preference, we found that ester-hydrolases with prominent promiscuity are not enantio-selective. This last issue is of critical relevance as industry is interested in finding promiscuous enzymes that can be applied in multiple conversions, but the fact that those enzymes cannot be used in enantio-selective conversion limit their applicability. For this reason, enzyme engineering is expected to be needed to improve stereo-specificity in promiscuous ester-hydrolases. This was not the case of transaminases, as we found some of them capable of converting a large number of substrates while at the same time being (S)-specific. However, it is to note that none of the transaminases were (R)-specific.

Fourth, this Thesis has also delivered a number of microbial ester-hydrolases and transaminases that may compete with the best commercial preparations in terms of substrate spectra and selectivity.

Although it is difficult to predict *a priori* whether a new enzyme could replace the best commercial prototypes, it is at least possible to evaluate whether some of their properties are superior or not. Here, through comparing the substrate spectra with a small set of best commercial preparations, we were able to identify a set of enzymes with relevant properties. Among the 25 selected candidates (see above), we should point at least 3 ester-hydrolases with the capacity to convert a higher number of substrates, and 1 transaminase with broad substrate profile and a different selectivity compared to previously known similar enzymes.

The idea was to start with the ester-hydrolase with proven top "substrate promiscuity" and use this as a lead scaffold to engineer selectivity or increase substrate performance towards the desired reaction by applying two different and complementary engineering strategies, one being supramolecular engineering through immobilization, and one rational protein engineering.

Fifth, in relation to the protein engineering effort, this Thesis contributes to generating the Plurizyme concept: developing enzymes with additional (artificial) active sites with potentially improved catalytic properties.

In detail, this Thesis significantly challenges enzyme engineering by introducing, in a natural serine ester-hydrolase, a second hydrolytic active site. The original idea was to use the most "substrate promiscuous" but non enantio-specific ester-hydrolase as a lead scaffold to evaluate the possibility to increase its catalytic activity and enantio-selectivity by introducing an extra active site. Briefly, using the PELE software, a potential binding pocket was identified and successfully turned into a catalytic site by introducing a few mutations as shown in **Figure 24**. Although the expected increase in catalytic performance could not be obtained, the first enzyme with two separate (and active) catalytic sites was designed. This successful example may open multiple opportunities in relation to introducing in a natural/native enzyme different substrate specificities or selectivities, complementary biochemistry, different steps in cascade, etc. These possibilities are being currently evaluated (Nature Catalysis, 2018).

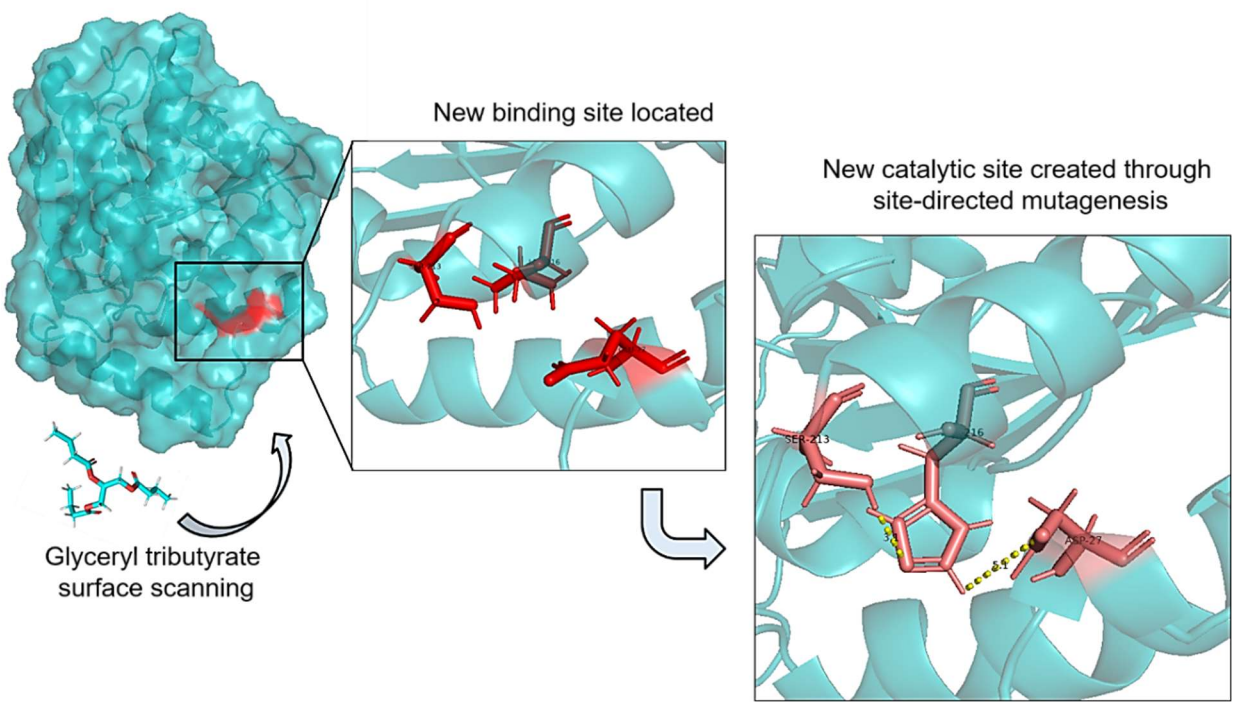


Figure 24. Schematic representation of workflow for creating a PluriZyme. First, using PELE a potentian place is identified locating a selected substrate. Second, by introducing appropriated mutations, this binding site is transformed into a catalytic site.

Sixth, this Thesis delves into the concept of Supramolecular Enzyme Engineering, which refers to changing enzyme properties by establishing a set of interactions at the surface of the enzyme through different immobilization strategies.

In detail, the original idea was to use the most "substrate promiscuous" but non "enantio-specific" ester-hydrolase as a lead scaffold to evaluate the possibility of increasing its enantio-selectivity through its immobilization by different techniques. We demonstrated that by immobilizing an enzyme in the interior of a mesoporous material or on a surface of micro-particles, one can differentially tune substrate selectivity and endow the enzyme with enantio-selective properties. Although this was expected, the most interesting outcome was that when the enzyme is immobilized inside pores with dimensions close to that of the enzyme, then it becomes enantio-specific. This did not happen when the pore dimension was significantly higher than that of the enzyme, and thus it has more movement freedom. Similarly, even when the enzyme was immobilized on a surface, the manner through which the enzyme was linked influenced in different ways the substrate specificity and enantio-specificity.

The results of the immobilization tests are of interest in the way that in this Thesis we found that promiscuous ester-hydrolases are not enantio-specific. However, by immobilization-based supramolecular enzyme engineering, promiscuous ester-hydrolases can be converted into enantio-specific ones, while still retaining a broad substrate spectrum. We anticipate that the flexibility constraints driven by the immobilization influence the differential access to the active site and positioning of substrates in it, thus promoting in different manner the specificity and selectivity. A scheme of the loss of flexibility through immobilization is shown in **Figure 25**.

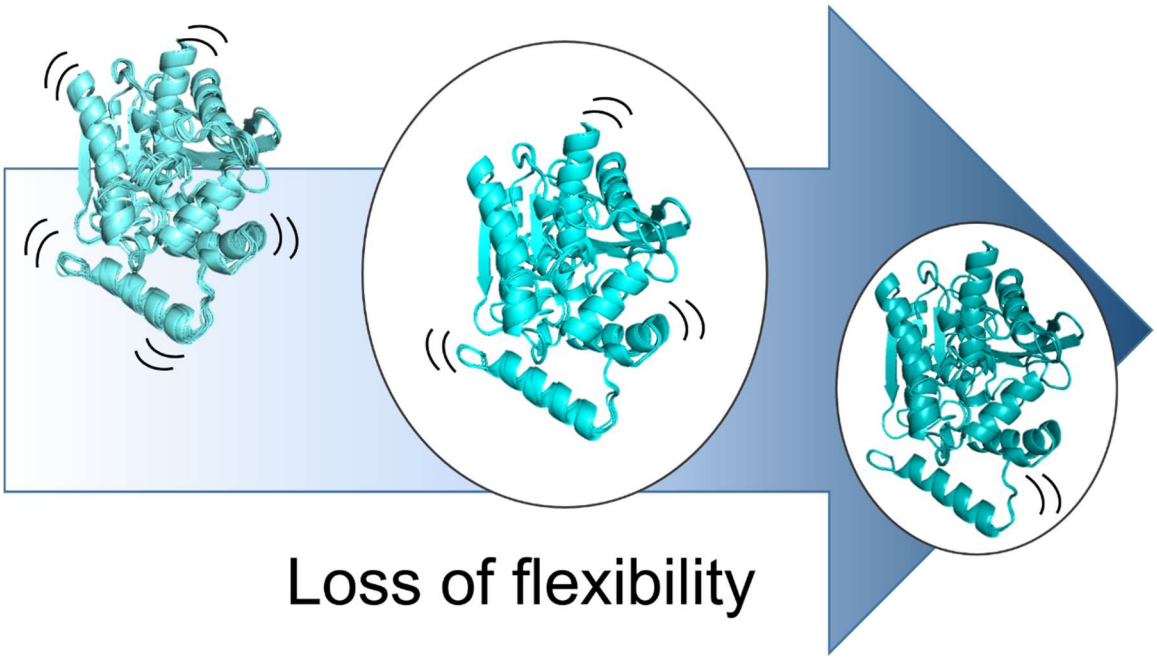


Figure 25. Schematic representation of the loss of flexibility of the structure of EH1 serine ester-hydrolase (PDB 5JD4) scaffold through immobilization inside a mesoporous material with a pore size close to enzyme size.

To sum up, through providing a large bio-catalytic dataset for an ample set of ester-hydrolases and transaminases, this Thesis has provided a sub-set of those with properties superior to best commercial prototypes, background knowledge that may open new research avenues for deepening into enzyme properties and their prediction, to get more out of the sequence space in regards of being able to predict whether a sequence will most likely encode an enzyme with broad substrate spectra, and finally background knowledge that help engineering biocatalysts with improved properties, particularly, being "substrate promiscuous" while at the same time being "enantio-specific".

CONCLUSIONS

1. **Genomic and metagenomic techniques are a powerful tool to access the high Nature's biodiversity to find novel enzymes with catalytic properties suitable for industrial processes.**
2. **This Thesis has contributed to the creation of one of the largest collection of highly diverse ester-hydrolases and class III ω-transaminases, 155 in total, for which biocatalytic data have been obtained by applying high-throughput assays with about 132 chemically and structurally diverse substrates.**
3. **Ester-hydrolases and class III ω-transaminases with broad substrate range and capable of converting large and recalcitrant substrates are more abundant than previously thought.**
4. **Promiscuous ester-hydrolases are not enantio-specific, but promiscuous class III ω-transaminases can be enantio-specific.**
5. **Enzyme properties such as level of substrate promiscuity and enantio-specificity could be inferred through the analysis of a sequence. In the particular case of ester-hydrolases this can be done by calculating the effective volume, and in case of class III ω-transaminases by evaluating the presence of a hairpin region close to a conserved Arginine residue and its orientation.**
6. **By differentially influencing the flexibility or rigidity constraints of an enzyme structure one can modify *a la carte* the stereo-specificity and substrate spectra to an extent higher than previously thought.**
7. **The possibility to create ester-hydrolases with more than one active site is feasible.**
8. **Native enzymes with properties superior to those of commercial prototypes can still be identified through genome and metagenome mining.**

CONCLUSIONES

1. **Las técnicas genómicas y metagenómicas son una herramienta poderosa que nos permite acceder a la gran biodiversidad presente en la naturaleza para encontrar enzimas con propiedades catalíticas apropiadas para los procesos industriales.**
2. **Esta Tesis ha contribuido a la creación de una de las colecciones más grandes de éster-hidrolasas y ω-transaminasas de la clase III ampliamente diversas, 155 en total, para las cuales se han obtenido datos biocatalíticos mediante la aplicación de ensayos masivos con alrededor de 132 sustratos química y estructuralmente distintos.**
3. **Las éster-hidrolasas y ω-transaminasas de clase III capaces de convertir sustratos voluminosos y recalcitrantes son más abundantes de lo que se pensaba hasta ahora.**
4. **Las éster-hidrolasas promiscuas no son enantio-específicas, pero las transaminasas promiscuas sí pueden ser enantio-específicas.**
5. **Las propiedades enzimáticas como el nivel de promiscuidad de sustrato y la especificidad y enantio-selectividad pueden ser inferidas mediante el análisis de secuencia. En el caso particular de las éster-hidrolasas esto puede ser realizado mediante el cálculo del volumen efectivo, y en el caso de las ω-transaminasas de la clase III mediante la presencia de una región horquilla cerca de una Arginina conservada y la orientación de esta última.**
6. **Modificando las restricciones de flexibilidad de una estructura enzimática se puede modificar a la carta la estéreo-especificidad y el perfil de sustratos aceptados hasta niveles antes desconocidos.**
7. **Es posible crear éster-hidrolasas con más de un centro activo.**
8. **Enzimas naturales con propiedades superiores a las de prototipos comercialmente disponibles pueden ser identificadas en genomas y metagenomas.**

REFERENCES

Adesioye, F. A., Makhalanyane, T. P., Vikram, S., Sewell, B. T., Schubert, W.-D., & Cowan, D. A. (2018). Structural characterization and directed evolution of a novel acetyl xylan esterase reveals thermostability determinants of the carbohydrate esterase 7 family. *Applied and Environmental Microbiology*, 84(8), e02695-17. https://doi.org/10.1128/AEM.02695-17

Albesa-Jové, D., & Guerin, M. E. (2016). The conformational plasticity of glycosyltransferases. *Current Opinion in Structural Biology*, 40, 23–32. https://doi.org/https://doi.org/10.1016/j.sbi.2016.07.007

Alcaide, M., Tornés, J., Stogios, P. J., Xu, X., Gertler, C., Di Leo, R., … Ferrer, M. (2013). Single residues dictate the co-evolution of dual esterases: MCP hydrolases from the α/β hydrolase family. *Biochemical Journal*, 454(1), 157– 166. https://doi.org/10.1042/BJ20130552

Aranda, J., Cerqueira, N. M. F. S. A., Fernandes, P. A., Roca, M., Tuñon, I., & Ramos, M. J. (2014). The catalytic mechanism of carboxylesterases: A computational study. *Biochemistry*, 53(36), 5820–5829. https://doi.org/10.1021/bi500934j

Arbige, M. V, & Pitcher, W. H. (1989). Industrial enzymology: A look towards the future. *Trends in Biotechnology*, 7(12), 330–335. https://doi.org/https://doi.org/10.1016/0167-7799(89)90032-2

Arnold, F. H. (2019). Innovation by evolution: Bringing new chemistry to life (Nobel lecture). *Angewandte Chemie International Edition*, 58(41), 14420–14426. https://doi.org/10.1002/anie.201907729

Arpigny, J. L., & Jaeger, K. E. (1999). Bacterial lipolytic enzymes: classification and properties. *The Biochemical Journal*, 343 Pt 1(Pt 1), 177–183. https://www.ncbi.nlm.nih.gov/pubmed/10493927

Aumala, V., Mollerup, F., Jurak, E., Blume, F., Karppi, J., Koistinen, A. E., … Master, E. R. (2019). Biocatalytic production of amino carbohydrates through oxidoreductase and transaminase cascades. *ChemSusChem,* 12(4), 848–857. https://doi.org/10.1002/cssc.201802580

Babtie, A., Tokuriki, N., & Hollfelder, F. (2010). What makes an enzyme promiscuous? *Current Opinion in Chemical Biology*, 14(2), 200–207. https://doi.org/https://doi.org/10.1016/j.cbpa.2009.11.028

Berenguer, J., & Ventosa, A. (2018). A brief reflection of International Microbiology's history and future direction. *International Microbiology*, 21(1), 1–2. https://doi.org/10.1007/s10123-018-0013-2

Berezovsky, I. N., Guarnera, E., Zheng, Z., Eisenhaber, B., & Eisenhaber, F. (2017). Protein function machinery: From basic structural units to modulation of activity. *Current Opinion in Structural Biology*, 42, 67–74. https://doi.org/https://doi.org/10.1016/j.sbi.2016.10.021

Bhatia, S., & Goli, D. (2018). Introduction to pharmaceutical biotechnology, volume 1. Basic techniques and concepts. *IOP publishing*. https://doi.org/10.1088/978-0-7503-1299-8

Blomberg, R., Kries, H., Pinkas, D. M., Mittl, P. R. E., Grütter, M. G., Privett, H. K., … Hilvert, D. (2013). Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature*, 503, 418. https://doi.org/10.1038/nature12623

Bommarius, A. S. (2015). Biocatalysis: A status report. *Annual Review of Chemical and Biomolecular Engineering*, 6(1), 319–345. https://doi.org/10.1146/annurev-chembioeng-061114-123415

Bommarius, A. S., Blum, J. K., & Abrahamson, M. J. (2011). Status of protein engineering for biocatalysts: How to design an industrially useful biocatalyst. *Current Opinion in Chemical Biology*, 15(2), 194–200. https://doi.org/https://doi.org/10.1016/j.cbpa.2010.11.011

Bornscheuer, U. T., Huisman, G. W., Kazlauskas, R. J., Lutz, S., Moore, J. C., & Robins, K. (2012). Engineering the third wave of biocatalysis. *Nature*, 485(7397), 185–194. https://doi.org/10.1038/nature11117

Borrelli, K. W., Cossins, B., & Guallar, V. (2010). Exploring hierarchical refinement techniques for induced fit docking with protein and ligand flexibility. *Journal of Computational Chemistry*, 31(6), 1224–1235. https://doi.org/10.1002/jcc.21409

Borrelli, K. W., Vitalis, A., Alcantara, R., & Guallar, V. (2005). PELE: Protein energy landscape exploration. A novel Monte Carlo based technique. *Journal of Chemical Theory and Computation*, 1(6), 1304–1311. https://doi.org/10.1021/ct0501811

Carlson, H. A., Smith, R. D., Damm-Ganamet, K. L., Stuckey, J. A., Ahmed, A., Convery, M. A., … Dunbar, J. B. (2016). CSAR 2014: A benchmark exercise using unpublished data from pharma. *Journal of Chemical Information and Modeling*, 56(6), 1063–1077. https://doi.org/10.1021/acs.jcim.5b00523

Copley, S. D. (2003). Enzymes with extra talents: Moonlighting functions and catalytic promiscuity. *Current Opinion in Chemical Biology*, 7(2), 265–272. https://doi.org/https://doi.org/10.1016/S1367-5931(03)00032-2

Copley, S. D. (2015). An evolutionary biochemist's perspective on promiscuity. *Trends in Biochemical Sciences*, 40(2), 72–78. https://doi.org/10.1016/j.tibs.2014.12.004

Daiha, K. de G., Angeli, R., de Oliveira, S. D., & Almeida, R. V. (2015). Are lipases still important biocatalysts? A study of scientific publications and patents for technological forecasting. *Plos One*, 10(6), e0131624. https://doi.org/10.1371/journal.pone.0131624

Damborský, J., & Koca, J. (1999). Analysis of the reaction mechanism and substrate specificity of haloalkane dehalogenases by sequential and structural comparisons. *Protein Engineering*, 12, 989–998. https://doi.org/10.1093/protein/12.11.989

Datta, S., Christena, L. R., & Rajaram, Y. R. S. (2013). Enzyme immobilization: An overview on techniques and support materials. *3 Biotech*, 3(1), 1–9. https://doi.org/10.1007/s13205-012-0071-7

Deponte, M., Sturm, N., Mittler, S., Harner, M., Mack, H., & Becker, K. (2007). Allosteric coupling of two different functional active sites in monomeric *Plasmodium falciparum* glyoxalase I. *Journal of Biological Chemistry* , 282(39), 28419–28430. https://doi.org/10.1074/jbc.M703271200

Devamani, T., Rauwerdink, A. M., Lunzer, M., Jones, B. J., Mooney, J. L., Tan, M. A. O., … Kazlauskas, R. J. (2016). Catalytic promiscuity of ancestral esterases and hydroxynitrile lyases. *Journal of the American Chemical Society*, 138(3), 1046–1056. https://doi.org/10.1021/jacs.5b12209

Diniz, W. J. S., & Canduri, F. (2017). Bioinformatics: An overview and its applications. *Genetics and Molecular Research*, 16(1), 1–21. https://doi.org/10.4238/gmr16019645

Ereky, K. (1919). Biotechnologie der Fleisch-, Fett-, und Milcherzeugung im landwirtschaftlichen Grossbetriebe : für naturwissenschaftlich gebildete Landwirte verfasst. *Berlin: P. Parey*.

Espaillat, A., Carrasco-López, C., Bernardo-García, N., Pietrosemoli, N., Otero, L. H., Álvarez, L., … Cava, F. (2014). Structural basis for the broad specificity of a new family of amino-acid racemases. *Acta*

Crystallographica. *Section D, Biological Crystallography*, 70(Pt 1), 79–90. https://doi.org/10.1107/S1399004713024838

Ferrer, M., Bargiela, R., Martínez-Martínez, M., Mir, J., Koch, R., Golyshina, O. V, & Golyshin, P. N. (2015). Biodiversity for biocatalysis: A review of the α/β-hydrolase fold superfamily of esterases-lipases discovered in metagenomes. *Biocatalysis and Biotransformation*, 33(5–6), 235–249. https://doi.org/10.3109/10242422.2016.1151416

Ferrer, M., Martínez-Martínez, M., Bargiela, R., Streit, W. R., Golyshina, O. V, & Golyshin, P. N. (2016). Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microbial Biotechnology*, 9(1), 22–34. https://doi.org/10.1111/1751-7915.12309

Frickel, E.-M., Jemth, P., Widersten, M., & Mannervik, B. (2001). Yeast glyoxalase I is a monomeric enzyme with two active sites. *Journal of Biological Chemistry*, 276(3), 1845–1849.

Gajda, A. D., Pawełczak, M., & Drag, M. (2012). Substrate specificity screening of oat (*Avena sativa*) seeds aminopeptidase demonstrate unusually broad tolerance in S1 pocket. *Plant Physiology and Biochemistry*, 54, 6–9. https://doi.org/https://doi.org/10.1016/j.plaphy.2012.02.006

Hajighasemi, M., Nocek, B. P., Tchigvintsev, A., Brown, G., Flick, R., Xu, X., … Yakunin, A. F. (2016). Biochemical and structural insights into enzymatic depolymerization of polylactic acid and other polyesters by microbial carboxylesterases. *Biomacromolecules*, 17(6), 2027–2039. https://doi.org/10.1021/acs.biomac.6b00223

Han, S.-W., Park, E.-S., Dong, J.-Y., & Shin, J.-S. (2015). Active-site engineering of ω-transaminase for production of unnatural amino acids carrying a side chain bulkier than an ethyl substituent. *Applied and Environmental Microbiology*, 81(20), 6994–7002. https://doi.org/10.1128/AEM.01533-15

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology*, 5(10), 245–249. https://doi.org/https://doi.org/10.1016/S1074-5521(98)90108-9

Hernández-Ortega, A., Borrelli, K., Ferreira, P., Medina, M., Martínez, A. T., & Guallar, V. (2011). Substrate diffusion and oxidation in GMC oxidoreductases: an experimental and computational study on fungal aryl-alcohol oxidase. *Biochemical Journal*, 436(2), 341–350. https://doi.org/10.1042/BJ20102090

Hodgson, J. (2019). Biotech's baby boom. *Nature Biotechnology*, 37(5), 502–512. https://doi.org/10.1038/s41587-019-0112-4

Höhne, M., Schätzle, S., Jochens, H., Robins, K., & Bornscheuer, U. T. (2010). Rational assignment of key motifs for function guides *in silico* enzyme identification. *Nature Chemical Biology*, 6(11), 807.

Huang, H., Pandya, C., Liu, C., Al-Obaidi, N. F., Wang, M., Zheng, L., … Farelli, J. D. (2015). Panoramic view of a superfamily of phosphatases through substrate profiling. *Proceedings of the National Academy of Sciences*, 112(16), E1974-E1983. https://doi.org/10.1073/pnas.1423570112

Huang, R., Hippauf, F., Rohrbeck, D., Haustein, M., Wenke, K., Feike, J., … Barkman, T. J. (2012). Enzyme functional evolution through improved catalysis of ancestrally nonpreferred substrates. *Proceedings of the National Academy of Sciences*, 109(8), 2966 – 2971. https://doi.org/10.1073/pnas.1019605109

Hult, K., & Berglund, P. (2007). Enzyme promiscuity: mechanism and applications. *Trends in Biotechnology*, 25(5), 231–238. https://doi.org/https://doi.org/10.1016/j.tibtech.2007.03.002

Javed, S., Azeem, F., Hussain, S., Rasul, I., Siddique, M. H., Riaz, M., … Nadeem, H. (2018). Bacterial lipases: A review on purification and characterization. *Progress in Biophysics and Molecular Biology*, 132, 23–34. https://doi.org/https://doi.org/10.1016/j.pbiomolbio.2017.07.014

Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Röthlisberger, D., Zanghellini, A., … Baker, D. (2008). *De novo* computational design of retro-aldol enzymes. *Science (New York, N.Y.),* 319(5868), 1387–1391. https://doi.org/10.1126/science.1152692

Keller, D., Beloqui, A., Martínez-Martínez, M., Ferrer, M., & Delaittre, G. (2017). Nitrilotriacetic amine-functionalized polymeric core–shell nanoparticles as enzyme immobilization supports. *Biomacromolecules*, 18(9), 2777–2788. https://doi.org/10.1021/acs.biomac.7b00677

Khersonsky, O., Lipsh, R., Avizemer, Z., Ashani, Y., Goldsmith, M., Leader, H., … Fleishman, S. J. (2018). Automated design of efficient and functionally diverse enzyme repertoires. *Molecular Cell*, 72(1), 178-186.e5. https://doi.org/10.1016/j.molcel.2018.08.033

Kmunícek, J., Hynková, K., Jedlicka, T., Nagata, Y., Negri, A., Gago, F., … Damborský, J. (2005). Quantitative analysis of substrate specificity of haloalkane dehalogenase LinB from *Sphingomonas paucimobilis* UT26. *Biochemistry*, 44(9), 3390–3401. https://doi.org/10.1021/bi047912o

Koudelakova, T., Chovancova, E., Brezovsky, J., Monincova, M., Fortova, A., Jarkovsky, J., & Damborsky, J. (2011). Substrate specificity of haloalkane dehalogenases. *The Biochemical Journal*, 435(2), 345–354. https://doi.org/10.1042/bj20101405

Kunath, B. J., Bremges, A., Weimann, A., McHardy, A. C., & Pope, P. B. (2017). Metagenomics and CAZyme discovery BT - protein-carbohydrate interactions: methods and protocols. *D. W. Abbott & A. Lammerts van Bueren, Eds*. https://doi.org/10.1007/978-1-4939-6899-2_20

Kyrpides, N. C., Hugenholtz, P., Eisen, J. A., Woyke, T., Göker, M., Parker, C. T., … Klenk, H.-P. (2014). Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains. *PLOS Biology*, 12(8), e1001920. https://doi.org/10.1371/journal.pbio.1001920

Lecina, D., Gilabert, J. F., & Guallar, V. (2017). Adaptive simulations, towards interactive protein-ligand modeling. *Scientific Reports*, 7(1), 8466. https://doi.org/10.1038/s41598-017-08445-5

Li, A., & Shao, Z. (2014). Biochemical characterization of a haloalkane dehalogenase DadB from *Alcanivorax dieselolei* B-5. *Plos One*, 9(2), e89144. https://doi.org/10.1371/journal.pone.0089144

Liu, C., Chen, R., Sera, F., Vicedo-Cabrera, A. M., Guo, Y., Tong, S.,…, Kan, H. (2019). Ambient particulate air pollution and daily mortality in 652 cities. *New England Journal of Medicine*, 381 (2019), pp. 705-715. https://doi.org/10.1056/NEJMoa1817364

London, N., Farelli, J. D., Brown, S. D., Liu, C., Huang, H., Korczynska, M., … Shoichet, B. K. (2015). Covalent docking predicts substrates for haloalkanoate dehalogenase superfamily phosphatases. *Biochemistry*, 54(2), 528–537. https://doi.org/10.1021/bi501140k

Market, V. F. (2016). Global opportunity analysis and industry forecast, 2017-2023 *Allied Market Research: Pune*. Maharashtra, India: August.

Martínez-Martínez, M., Coscolín, C., Santiago, G., Chow, J., Stogios, P. J., Bargiela, R., … Ferrer, M. (2018). Determinants and prediction of esterase substrate promiscuity patterns. *ACS Chemical Biology*, 13(1), 225–234. https://doi.org/10.1021/acschembio.7b00996

Mate, D. M., Rivera, N. R., Sanchez-Freire, E., Ayala, J. A., Berenguer, J., & Hidalgo, A. (2019). Thermostability enhancement of the *Pseudomonas fluorescens* esterase I by *in vivo* folding selection in *Thermus thermophilus*. *Biotechnology and Bioengineering*, 0(0). https://doi.org/10.1002/bit.27170

Montella, S., Ventorino, V., Lombard, V., Henrissat, B., Pepe, O., & Faraco, V. (2017). Discovery of genes coding for carbohydrate-active enzyme by metagenomic analysis of lignocellulosic biomasses. *Scientific Reports*, 7, 42623. https://doi.org/10.1038/srep42623

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B., & Worm, B. (2011). How many species are there on earth and in the ocean? *PLOS Biology*, 9(8), e1001127. https://doi.org/10.1371/journal.pbio.1001127

Nature Catalysis. (2018). On advances and challenges in biocatalysis. *Nature Catalysis*, 1(9), 635–636. https://doi.org/10.1038/s41929-018-0157-7

Nobeli, I., Favia, A., & Thornton, J. (2009). Protein promiscuity and its implications for biotechnology. *Nature Biotechnology*, 27, 157–167. https://doi.org/10.1038/nbt1519

Ozer, A., Ay Sal, F., Belduz, A. O., Kirci, H., & Canakci, S. (2019). Use of feruloyl esterase as laccase-mediator system in paper bleaching. *Applied Biochemistry and Biotechnology*. https://doi.org/10.1007/s12010-019-03122-x

Panda, T., & Gowrishankar, B. S. (2005). Production and applications of esterases. *Applied Microbiology and Biotechnology*, 67(2), 160–169. https://doi.org/10.1007/s00253-004-1840-y

Park, E.-S., Kim, M., & Shin, J.-S. (2012). Molecular determinants for substrate selectivity of ω-transaminases. *Applied Microbiology and Biotechnology*, 93(6), 2425–2435. https://doi.org/10.1007/s00253-011-3584-9

Pellis, A., Cantone, S., Ebert, C., & Gardossi, L. (2018). Evolving biocatalysis to meet bioeconomy challenges and opportunities. *New Biotechnology*, 40, 154–169. https://doi.org/10.1016/j.nbt.2017.07.005

Price, D. R. G., & Wilson, A. C. C. (2014). A substrate ambiguous enzyme facilitates genome reduction in an intracellular symbiont. *BMC Biology*, 12(1), 110. https://doi.org/10.1186/s12915-014-0110-4

Purg, M., Pabis, A., Baier, F., Tokuriki, N., Jackson, C., & Kamerlin, S. C. L. (2016). Probing the mechanisms for the selectivity and promiscuity of methyl parathion hydrolase. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, 374(2080), 20160150. https://doi.org/10.1098/rsta.2016.0150

Puspita, I. D., Kamagata, Y., Tanaka, M., Asano, K., & Nakatsu, C. H. (2012). Are uncultivated bacteria really uncultivable? *Microbes and Environments*, 27(4), 356–366. https://doi.org/10.1264/jsme2.me12092

R.K. Pachauri, L. A. M. & C. W. T. (2014). Climate change 2014 synthesis report summary chapter for policymakers. *Ipcc,* 151. https://doi.org/10.1017/CBO9781107415324

Ramírez-Escudero, M., del Pozo, M. V, Marín-Navarro, J., González, B., Golyshin, P. N., Polaina, J., … Sanz-Aparicio, J. (2016). Structural and functional characterization of a ruminal β-glycosidase defines a novel subfamily of glycoside hydrolase family 3 with permuted domain topology. *Journal of Biological Chemistry*, 291(46), 24200–24214. https://doi.org/10.1074/jbc.M116.747527

Rathi, P. C., Fulton, A., Jaeger, K.-E., & Gohlke, H. (2016). Application of rigidity theory to the thermostabilization of lipase A from *Bacillus subtilis*. *PLoS Computational Biology*, 12(3), e1004754–e1004754. https://doi.org/10.1371/journal.pcbi.1004754

Reyes-Duarte, D., Coscolín, C., Martínez-Martínez, M., Ferrer, M., & García-Arellano, H. (2018). Functional-based screening methods for detecting esterase and lipase activity against multiple substrates lipases and phospholipases: methods and protocols. *G. Sandoval, Ed.* https://doi.org/10.1007/978-1-4939-8672-9_4

Richter, F., Blomberg, R., Khare, S. D., Kiss, G., Kuzin, A. P., Smith, A. J. T., … Baker, D. (2012). Computational design of catalytic dyads and oxyanion holes for ester hydrolysis. *Journal of the American Chemical Society*, 134(39), 16197–16206. https://doi.org/10.1021/ja3037367

Rodrigues, C. J. C., Pereira, R. F. S., Fernandes, P., Cabral, J. M. S., & de Carvalho, C. C. C. R. (2017). Cultivation-based strategies to find efficient marine biocatalysts. *Biotechnology Journal*, 12(7), 1700036. https://doi.org/10.1002/biot.201700036

Rodrigues, C. J. C., Sanches, J. M. R., & de Carvalho, C. C. C. R. (2019). Determining transaminase activity in bacterial libraries by time-lapse imaging. *Chemical* Communications, 55, 13538-13541 https://doi.org/10.1039/C9CC07507K

Romano, D., Bonomi, F., de Mattos, M. C., de Sousa Fonseca, T., de Oliveira, M. da C. F., & Molinari, F. (2015). Esterases as stereoselective biocatalysts. *Biotechnology Advances*, 33(5), 547–565. https://doi.org/https://doi.org/10.1016/j.biotechadv.2015.01.006

Rufo, C. M., Moroz, Y. S., Moroz, O. V, Stöhr, J., Smith, T. A., Hu, X., … Korendovych, I. V. (2014). Short peptides self-assemble to produce catalytic amyloids. *Nature Chemistry*, 6(4), 303–309. https://doi.org/10.1038/nchem.1894

Santiago, G., de Salas, F., Lucas, M. F., Monza, E., Acebes, S., Martinez, Á. T., … Guallar, V. (2016). Computer-aided laccase engineering: toward biological oxidation of arylamines. *ACS Catalysis*, 6(8), 5415–5423. https://doi.org/10.1021/acscatal.6b01460

Sawaya, D., & Internacional, P. (2010). La evolución de la bioeconomía hasta 2030: diseño de una agenda política. *Monográfico. Nota d'economia 97-98. 3er cuatrimestre* (2010) 11–32.

Sayali, K., Sadichha, P., & Surekha, S. (2013). Microbial esterases: An overview. *International Journal of Current Microbiology and Applied Sciences*, 2(7), 135–146.

Sayer, C., Martinez-Torres, R. J., Richter, N., Isupov, M. N., Hailes, H. C., Littlechild, J. A., & Ward, J. M. (2014). The substrate specificity, enantioselectivity and structure of the (R)-selective amine: pyruvate transaminase from *Nectria haematococca*. *The FEBS Journal*, 281(9), 2240–2253. https://doi.org/10.1111/febs.12778

Shan, Y., Kim, E. T., Eastwood, M. P., Dror, R. O., Seeliger, M. A., & Shaw, D. E. (2011). How does a drug molecule find its target binding site? *Journal of the American Chemical Society*, 133(24), 9181–9183. https://doi.org/10.1021/ja202726y

Sheldon, R. A., & Brady, D. (2019). Broadening the scope of biocatalysis in sustainable organic synthesis. *ChemSusChem*, 12(13), 2859–2881. https://doi.org/10.1002/cssc.201900351

Simpson, P. J., Tantitadapitak, C., Reed, A. M., Mather, O. C., Bunce, C. M., White, S. A., & Ride, J. P. (2009). Characterization of two novel aldo–keto reductases from *Arabidopsis*: Expression patterns, broad substrate specificity, and an open active-site structure suggest a role in toxicant metabolism

following stress. *Journal of Molecular Biology*, 392(2), 465–480. https://doi.org/https://doi.org/10.1016/j.jmb.2009.07.023

Steffen-Munsberg, F., Vickers, C., Kohls, H., Land, H., Mallin, H., Nobili, A., … Bornscheuer, U. T. (2015). Bioinformatic analysis of a PLP-dependent enzyme superfamily suitable for biocatalytic applications. *Biotechnology Advances*, 33(5), 566–604. https://doi.org/https://doi.org/10.1016/j.biotechadv.2014.12.012

Sukul, P., Schäkermann, S., Bandow, J. E., Kusnezowa, A., Nowrousian, M., & Leichert, L. I. (2017). Simple discovery of bacterial biocatalysts from environmental samples through functional metaproteomics. *Microbiome*, 5(1), 28. https://doi.org/10.1186/s40168-017-0247-9

Suplatov, D., & Švedas, V. (2015). Study of functional and allosteric sites in protein superfamilies. *Acta Naturae*, 7(4), 34–45. https://www.ncbi.nlm.nih.gov/pubmed/26798490

Teilum, K., Olsen, J. G., & Kragelund, B. B. (2009). Functional aspects of protein flexibility. *Cellular and Molecular Life Sciences*, 66(14), 2231–2247. https://doi.org/10.1007/s00018-009-0014-6

Toscano, M. D., Woycechowsky, K. J., & Hilvert, D. (2007). Minimalist active-site redesign: Teaching old enzymes new tricks. *Angewandte Chemie International Edition*, 46(18), 3212–3236. https://doi.org/10.1002/anie.200604205

Turner, N. J., & Truppo, M. D. (2013). Biocatalysis enters a new era. *Current Opinion in Chemical Biology*, 17(2), 212–214. https://doi.org/https://doi.org/10.1016/j.cbpa.2013.02.026

Ufarté, L., Laville, É., Duquesne, S., & Potocki-Veronese, G. (2015). Metagenomics for the discovery of pollutant degrading enzymes. *Biotechnology Advances*, 33(8), 1845–1854. https://doi.org/https://doi.org/10.1016/j.biotechadv.2015.10.009

Valdes, C., Stebliankin, V., & Narasimhan, G. (2019). Large scale microbiome profiling in the cloud. *Bioinformatics*, 35(14), i13–i22. https://doi.org/10.1093/bioinformatics/btz356

Welborn, V. V., Ruiz Pestana, L., & Head-Gordon, T. (2018). Computational optimization of electric fields for better catalysis design. *Nature Catalysis*, 1(9), 649–655. https://doi.org/10.1038/s41929-018-0109-2

Wilding, M., Peat, T. S., Newman, J., & Scott, C. (2016). A β-alanine catabolism pathway containing a highly promiscuous ω-transaminase in the 12-aminododecanate-degrading *Pseudomonas sp*. strain AAC. *Applied and Environmental Microbiology*, 82(13), 3846 – 3856. https://doi.org/10.1128/AEM.00665-16

Wilson, M. C., & Piel, J. (2013). Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. *Chemistry & Biology*, 20(5), 636–647. https://doi.org/https://doi.org/10.1016/j.chembiol.2013.04.011

World Health Organization (WHO). (2017). Evolution of WHO air quality guidelines. *In Copenhagen: WHO Regional Office for Europe*.

Yang, Y., Lin, J., & Wei, D. (2016). Heterologous overexpression and biochemical characterization of a nitroreductase from *Gluconobacter oxydans* 621H. *Molecular Biotechnology*, 58(6), 428–440. https://doi.org/10.1007/s12033-016-9942-1

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K.-H., … Rosselló-Móra, R. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology*, 12, 635. https://doi.org/10.1038/nrmicro3330

Ye, S. H., Siddle, K. J., Park, D. J., & Sabeti, P. C. (2019). Benchmarking metagenomics tools for taxonomic classification. *Cell*, 178(4), 779–794. https://doi.org/https://doi.org/10.1016/j.cell.2019.07.010

Yip, S. H.-C., & Matsumura, I. (2013). Substrate ambiguous enzymes within the *Escherichia coli* proteome offer different evolutionary solutions to the same problem. *Molecular Biology and Evolution*, 30(9), 2001–2012. https://doi.org/10.1093/molbev/mst105

Yoshida, K., & Komae, K. (2006). A rice family 9 glycoside hydrolase isozyme with broad substrate specificity for hemicelluloses in type II cell walls. *Plant and Cell Physiology*, 47(11), 1541–1554. https://doi.org/10.1093/pcp/pcl020